

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 20:58:35 ; Search time 5709 Seconds
(without alignments)
16682.013 Million cell updates/sec

Title: US-09-806-767-1
Perfect score: 2328
Sequence: 1 atggacccaacacattttt.....taaccttgagactaactga 2328

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 segs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.:

1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rod.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sy.*
39: em_hgt_hum.*
40: em_hgt_mus.*
41: em_hgt_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2328	100.0	78259	8	P20D22
2	691.2	29.7	150503	8	AC002411 Arabidops
3	482	20.7	2305	8	AC122149 Oryza sat
4	482	20.7	2618	8	AY142612 Arabidops
5	482	20.7	2686	8	AF360285 Arabidops
6	482	20.7	72698	8	AF065639 Arabidops
7	481.8	20.7	2265	6	AX506370 Sequence
8	481.8	20.7	2296	8	AY142613 Arabidops
9	481.8	20.7	2541	8	AY035090 Arabidops
10	480.4	20.6	2274	8	BT001082 Arabidops
11	480.4	20.6	2520	8	AY091773 Arabidops
12	456.6	19.6	2423	8	ATARA12
13	447.4	19.2	86212	8	AY080831 Arabidops
14	447.4	19.2	2549	8	AB010074 Arabidops
15	424.4	18.2	2525	8	AF360129 Arabidops
16	424.4	18.2	83906	8	AB022220 Arabidops
17	417	17.9	2638	8	AY084387 Arabidops
18	410.4	17.6	2645	8	AF436834 Arabidops
19	407.6	17.5	2650	8	LESBT2
20	407.6	17.5	5197	8	LES6379
21	398.2	17.1	2295	6	AX506491 Sequence
22	398.2	17.1	2326	8	AY096357 Arabidops
23	398.2	17.1	2665	8	AY074375 Arabidops
24	398.2	17.1	80346	8	ATM4E13
25	398.2	17.1	103150	8	ATF1111
26	398.2	17.1	195165	8	ATCHRIV82
27	397.2	17.1	2689	8	LESBT1
28	397.2	17.1	5089	8	LES6378
29	377.8	16.2	121160	8	AC005970 Arabidops
30	376.8	16.2	2465	8	BT002840 Arabidops
31	374.6	16.1	2152	8	PEU457185 Populus e
32	365.4	15.7	154863	2	OSJN00071 Arabidops
33	359	15.4	199414	2	AC124955 Medicago
34	354.6	15.2	128017	8	AC084282 Oryza sat
35	344.2	14.8	73859	8	AP000600 Arabidops
36	343.8	14.8	113049	8	AC134239 Oryza sat
37	330.4	14.2	47295	8	F22M8
38	312.8	13.4	3173	8	LES17278 Arabidops
39	312	13.4	2680	8	LEAJ5172 Arabidops
40	311.4	13.4	3540	8	LES18932 Arabidops
41	311.2	13.4	2849	8	LEAJ5173 Arabidops
42	311	13.4	2407	8	LESUBTILI
43	311	13.4	3076	8	LES17276 Arabidops
44	307.6	13.2	2847	8	LEAJ517 Arabidops
45	303.6	13.0	3975	8	LES18931 Arabidops

ALIGNMENTS

RESULT 1

F20D22

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

F20D22 78259 bp DNA linear PIN 20-MAY-1998
Arabidopsis thaliana chromosome 1 BAC F20D22 sequence, complete
sequence.

AC002411 GI:2570223

HTG.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 78259)

Vysotskaia.V.S., Osborne,B.I., Schwartz,J.R., Toriumi,M., Kwan,A.,

102(6)

QY	304	GCAGTGAGACCTGACCACTGTTCTTTCAGGTTCAACACACTTACTCTTACAGTTCTTTGGGA	363	Db	87746	GGCTACAGGAGGCGCTGAGCTCAAGAACTACGTCAGCTCCACGCGCGCGCTGGCG	87805
Db	86669	TCGGTTCGGGCGGACCGCGGGTTGAGCTGCACACGACGTACTCGTACAGTTCTTTGGG	86728	QY	1438	CGGATAATTTTGTGTGTACCGTGTATTTGGGAGGTCACGAGCACCGAGGTGGCTCAGTTT	1497
QY	364	CTCAGCGTTTGTGAACACTCCGGTGTATGGTCTAAATCTCCGTTTGGTCAAGCACAAAT	423	Db	87806	AGGATAGTGTTCGGCGGACCGGATTTGGCGGCGACGCGCGCGCGGTGGCGTGTTC	87865
Db	86729	CTCAGC---TTCGTCCGACCGGCGCGTGGCGCGTCCGGGTACGGCGCGCACGATC	86785	QY	1498	TCAGCTCCAGGACCGAGTTAGCCAATCTTTCCGATACTTAAACCGGATATGATTGCTCCG	1557
QY	424	ATCCGCGTGTGTGATATCGAGTTTGGCTGAAAGTCTCTAGTCTTACGATACCGGAATG	483	Db	87866	TCGCGCGCGCGCGAGCTGACGNAACCGTCTGGTCTCAAGCCCGAGTGGTCCGCCCC	87925
Db	86786	ATCGCGTGTCTCGACACCGCGTGTGGCGGAGAAATCCAGCTTCGATGACCGCGGATG	86845	QY	1558	GGAGTCAATATCATTTGCGGCTTGGCTCAAAATCTAGGACCAACCGGACTCTCTTATGAT	1617
QY	484	CTTTTCGATTCCACGGAATGGAAGAGATTTGCCAAGAGGAAAGTTTCAGTTCTTCG	543	Db	87926	GGCGTGAACATCATCGCGCTGGCCCGCAACCTGGGACCGTGGGGCTGGAGGCGCAC	87985
Db	86846	CCGCGGTGCGCGCGCGTGGCAGGCGGTGGCCAGGAGGTGAGCAATTTCAACGCCACC	86905	QY	1618	TCAGAAAGAGTTAACTTCACTGTATGTCAAGAACTTCAATGTCTTGTCCACATGTTAGC	1677
QY	544	AGCTGTAAACCGGAAGCTAATCCGTTGTAGATTCTTTCATCAGAGACACCGGTGCGTAAAT	603	Db	87986	GCCCGCGCGTCCGATTTCAAGCTGTCTCGGGGACATCAATGGCGTCCCGCAGCTGAGC	88045
Db	86906	AATTCGAAATCGAAGCTCATCGCGCGGGTTCATCTCAAGGGCACCGGCCAACTAC	86965	QY	1678	GGAACTACTGCTCTTATCCGCTCTGCATACCCGAACTGCTCTCCAGCTGCAATCAAAATCC	1737
QY	604	TCACCAGAGGAATCACCACATCCCTC---GTGAATACATTTCCGCAAGAGATTTCAACG	660	Db	88046	GGCATCGCGCGCGCTCATCAGGTCCGCGCACCGCTCGTGAGCCCGCGCATGGTGGGTTCC	88105
Db	86966	CCGACCAACCGTCCGACCGCGTGTGCTGATGAGTACGTGTGCGCGGGACGCGCAC	87025	QY	1738	GCATTTGATCAACAACAGCGGATTTGTACGATCGTCAAGGGAAGAGATAAAGATGTTAAC	1797
QY	661	GGACACGGGAATCACACCGCTCAACAGTTGGTGGATCTCTGTTTCGATGGCGAATGTT	720	Db	88106	GGCATCATGACGACCGCGCGAGCTGACGACCGCGGAGGCAAGCCGATCATGACGGGAAC	88165
Db	87026	GGGACACGGGACGACACCGCTCGACCGCGCGCGCGCGCGCGTGGCGCGCGCGAGCGTC	87085	QY	1798	AAACCA---GCCGCTGTGTTTCGATTTGGAGCAGGCGATGTGAATCCGCAAAAGCGGATA	1854
QY	721	CTTGGCAATGAGCTGGTGTGGCTCGTGGGATGGCTCTCGAGCTCACATGTCAGTCTAT	780	Db	88166	GGCGGCAAGGCGACGCGTACGCCATGGCGCGCGGCGACGTAACCCCGCGAGGCGCGTC	88225
Db	87086	CTGGCGGTGCGGGCGCGGACGCGCGCGCGCGTGGCGCGCGCGCGCGCGCGCGCTAC	87145	QY	1855	AACCGCGGATGTTTTCACAACTTCAACAGTGGATTAACAATTAACCTCTGCACTCTT	1914
QY	781	AAAGTCTCTGTTCAATGTTTGTACAGCTCTGACATTTCTAGCAGCTATAGATGTAGCG	840	Db	88226	GAACCGGCGCTCTCTAGACATACACCGCGCGCGCTACGTCACCCACTCTGCAACCTC	88285
Db	87146	AAAGTGTCTGTTCAACGGGTGCTACAGCTCCGACATCTCTCGCGGGATGGACGCGCC	87205	QY	1915	GGATTCAACAAGATCAGATATTTAGCAATCACTATAAGAAAGCTGAGGTGCAATGGAATA	1974
QY	841	ATTCAAGATAAAGTCGATGTTCTTTTCGCTTTCCCTTGGCGGTTTCCCTATTCCTTTGAT	900	Db	88286	GGGTACACCCACATGGAGATCTTCAAGATCACCCAGCTGGGCTCAACTGACCCCGGTG	88345
Db	87206	GTGCGCGAGCGGCTCGAGTGTGCTGCTCTCCCTCGCGCGCTTCCCATCCCGCTCTTC	87265	QY	1975	TTCCGGAATAAACCCCGGTTTATGTTCTCAATTAACCGCTCGATAGCCGTGATTTTCAACGT	2034
QY	901	GATGACACATCGCCATTTGGAACATTCGAGCGCATGGAACCGGTATATCTGTAATCTGT	960	Db	88346	CTGAGCGGAACCGCGCTTCAGCTCAACTACCCATCATATCGTGGCGTTCAAGACG	88405
Db	87266	GAGGACAGCATCGCCATCGGAGCTTCGCTGCCACGACGCAATGCGCTCTCGCTGTGTC	87325	QY	2035	GGCAAGACTACGAGATGATACAGGCGTGTCACTAAGTGGGAGTCTTAACCTGATA	2094
QY	961	GCAGCTGTATCAACCGTTCATCGAAGAGCTGTGTGCAACACAGCTCTCTGGGTCTCA	1020	Db	88406	AACACACCTCGCGCGTCTCCAGAGGACGCTCACCAACGTCGCGCACCGCAACTCGACG	88465
Db	87326	GCCGCGGGAAACACGGGCGCATCGCGGAGCTCGGTGCGCAACAGGCGCCATGSGTATC	87385	QY	2095	TACTCAGTGAATGTCAAGCTCCAGGCGGATCAAGTTTATTTGTCATCTTAAGAGACTT	2154
QY	1021	ACCATTTGCGCAGGACACGCTTGTATCGAAGATTTCCCGCTGTGGTCAAGTTAGCCAACGGA	1080	Db	88466	TACCCGCGCAGTCCCGCGCGCGCACCGCGTCAAGGTTGCGCGGCTGTCGCGCGGACGCTG	88525
Db	87386	ACCGTCCGCGCGGACGCTGGACCGCGGATTCGCGGCTATGTCGCGCTCGGGAACGCG	87445	RESULT 3			
QY	1081	AAGCTTCTATGAGAGATCAATGTTATCCGGGAAA---GGTATAAGAAATGCCGGAGA	1137	AY142612	Arabidopsis thaliana putative subtilisin serine protease ARA12	2305 bp	linear
Db	87446	CGGATCTGTACCGCGAGTCCATGTTCCCGGGAAGTCGATTTGAAATAATGGCGGGAAG	87505	LOCUS	AY142612		
QY	1138	GAGTTGAGGTGATTTACGTACAGAGAGAGATAAGGAAGTGTGTTCTGTTGAGAGG	1197	DEFINITION	AY142612	GI:23296831	
Db	87506	GAGCTGAGCTGTGTACCGCGGAGTGGGACCGCGGAGAGATGACTGTCATCAAGGG	87565	ACCESSION	AY142612		
QY	1198	TCACTTCCAGAGAGAAATCCGAGCAAAATGTTGATTTGTGATCGCGGAGTCAATGGA	1257	VERSION	AY142612.1		
Db	87566	GCCTGTCTGCGCGCACTGTCCCGGAAAGATGTTGTTGTGACCGTGGCATCACCGCG	87625	KEYWORDS	FLI CDNA		
QY	1258	AGATCGGAGAAAGGAGCGGTTAAAGAGCTGGAGAGTTGCAATGATCTTAGCCAAAT	1317	SOURCE	Arabidopsis thaliana (thale cress)		
Db	87626	AGGGCTGACAAAGGCGAGCGAGTAAACCAAGCAGCGCGCGCGCATGATTCTCGCAAT	87685	ORGANISM	Arabidopsis thaliana		
QY	1318	ACAGATCAACCAAGAGAAATCTATGAGCTTATCTCTTACAGCTACATGATT	1377	REFERENCE	1 (bases 1 to 2305)		
Db	87686	TCCAGATAAACCAGAGAGGACTCCGTCGATGTCATGTCTCGCGTCGACGCTCATC	87745	AUTHORS	Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Huan, V.W., Lee, J.M., Quach, H.L., Tang, C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R.,		
QY	1378	GGTTACACTGAGTCAGTCTTCTGAGGCTTATGTTATGCCACCGTGAACCAAGGCG	1437				

TITLE
JOURNAL
REFERENCE
AUTHORS

Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
Arabidopsis Open Reading Frame (ORF) Clones
Unpublished
2 (bases 1 to 2305)
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
Direct Submission
Submitted (16-AUG-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

TITLE
JOURNAL
COMMENT

The Salk, Stanford, PGEN (SSP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RAPL cDNAs: Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yu, G., Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W., Ecker, J.R. and Theologis, A.
Yamada, K. (SSP/PGEN) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEN) contributed equally to this work as PIs.
Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.
Location/Qualifiers
1. .2305
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="U09277"
/note="This clone is in PUNI 51.
ecotype: Columbia"
1. .2305
/gene="At5g67360"
1. .2274
/gene="At5g67360"
/note="contains subtilase domains PF00082 and a PA domain PF0225"
/codon_start=1
/evidence=experimental
/product="putative subtilisin serine protease ARA12"
/protein_id="AA13181.1"
/db_xref="GI:23296832"
/translation="MSRSFSSSTAFFILLCLGFCVSSSSDQGYIVHAKSQMPSS
FDLHNSWYDSLRISDSAEILTYENAIHGFSTLTFQEEADSLMTQPGVIVLPSHR
YELHTTRPLFLGIDHTADLFPFAGSVGVVGLDTGVWPKSKSYSDGFGPIPS
WKGCEAGTNTASLCNKRKLIGARFARGVSTWGPIDESKRSRPRDDGHTHTSS
TAAGSVBGSILGVASTAGRNAPRVARVYKVCWLGCFSSDILAIKADIVNV
VLMSLGGMSDYRDGNFTGVAIFAFAMERGIUVCSAGNAPFSSSLNAPWITTVGA
GTLDRDFALAILGNKRNFTGSLFKGELPKLLPFIYAGNASNATNGMLTGLI
PEKVKGTVMCDGNARVQKGVVKAAGVGMILANTAANGELVADADLHPATTIV
EKAGDIIRHYVTDTPNPASISILGVVGVKSPVAAPFSRGPNSITPNILKPLDIA
PGVNLIAWTAGPTGLASDRRVEFNIISGTSMSCPHVSGLAALLKSVHPWSPAA
IRSLAMTAYTKDKGLDIAATGKPSYVADLNYPFVAVNDGVGAYKTRIVTSV
GFLCALNTYSPQIRSVRNRYTCDPSKSYVADLNYPFVAVNDGVGAYKTRIVTSV
CGAGTYSVKVTSETTGKVISVEPAVLNFKEANRKSSTVTFTVDSKSPSGNSPFGSIE
2275..2305
/gene="At5g67360"

BASE COUNT	547 a	567 c	580 g	611 t
ORIGIN	Query Match Best Local Similarity 20.7%; Score 482; DB 8; Length 2305; Matches 974; Conservative 56.2%; Pred. No. 7.5e-126; 0; Mismatches 740; Indels 18; Gaps 3;			
Oy	210	ACTTCTCTACTCTATGGCTCTCGGATGAAGGATTTGCTGCAGTTCAGTCAATCAGA	269	
Db	192	ATTGCTCTACACTTACGAGAACCGGATTCATGGAATCTCAACAGGCTACTCAAGA	251	
Oy	270	AGCGAGATACAGATATTCACCTGAAAGTTGTTCAGTGAGACCTGACCATTTCTCA	329	
Db	252	AGCGACTCGCTCATGACTCAACTGGTGTATCTCGTTTTACCGGAGCACCGTTAC	311	
Oy	330	GGTTTCAAAACCACTTACTCTTTACAAGTTCTTGGGACTCGACGGTTTGGAAACT	389	
Db	312	GCTACACACCACCTGCTACTCTCTCTCTCTCGGTCTCGACG--AACACACCCG	368	
Oy	390	ATGCTCTAAATCTCGGTTTGGTCAAGCACAATATTCGGGCTGCTTGATCTGAG	449	
Db	369	GTTCCTCTGAAGCCGGCTCTTACAGCAGCTGCTCGGAGTTCTCGATACCGGAG	428	
Oy	450	GCCTGAAAGTCTAGCTTTGACGATACCGGAATGCCCTTCGATTCCACGGAAAT	509	
Db	429	CCCCGAGAGTAAAGCTTACTCCGACGAGGATTCGGTCTCTATTCCTCTCTG	488	
Oy	510	GATTTGCCAAGAAGGAGAAAGTTTTCAGTTCTTTCAGGCTGTAAACGGAAGCT	569	
Db	489	CGGATGCGAGCCGCGAACCAATTTTACCGCTTCTCTGTAAACGGAATTAAT	548	
Oy	570	TAGATTCTTATCAGAGGACACCGTGTGCTTAATTCACAGAGGAATCACAACAT	629	
Db	549	AAGATTCTTCGCTCGTGGTTAC-----GAATCAACGATGGGACCAATCG	599	
Oy	630	TCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCGCT	689	
Db	600	TAAAGATCAAGATCTCTAGAGACGACGCGTTCAGGAACTCACCTCATCGAC	659	
Oy	690	TGGTGATCTCTGTTTCGATGGGGAATGTTCTTGGCAATAGAGCTGGTGGCTG	749	
Db	660	CGCTGGATCCGTCGTTGAAGGAGCTAGCTTATTAGGCTACGCTTTCAGGA	719	
Oy	750	GATGCTCTGGAGCTCATATGAGCTCTATAAAGTCTGTTGGTTCAATGTTTAC	809	
Db	720	TATGCTCCACGCGCTGCTGCTGTTTACAAAGTCTGTTGGCTCGGTGTTGTT	779	
Oy	810	CTCTGACATCTAGCAGCTATAGATGATAGATCAAGATAAAGTCGATGTTCT	869	
Db	780	CTCAGATATTTTAGCTGCAATCGAATAAGCCATCGCCATACGTCATCTCTG	839	
Oy	870	TTCCCTTGGCGTTTCCCTATTCTGATGATGACACAATCGCCATTGGAAATTC	929	
Db	840	GTCTCTCGCGGTGGAATGTCGGATTTATTAGACCGGTGTTGGATCGGAGCT	899	
Oy	930	AGCCATGGAACCGGCTATATCTGTAATCTGTCAGCTGGTAACACGCTCAAT	989	
Db	900	CGCCATGGAAGAGGGAATTTTAGTATCTTGTCTCAGCTGGTAATGTCGG	959	
Oy	990	CTCTGTTGCAACACAGCTCTCTGGGTCTCAACCATTCGGCAGCAGCTTGAT	1049	
Db	960	TAGTTTATCAACAGTAGTCTCATGGATCAACACTGTGCTGCTGCTAGATCG	1019	
Oy	1050	ATTTCCCGCTGTGTCAGATTAGCCACCGAAAGCTTCTTATGAGAGTCAAT	1109	
Db	1020	TTTTCCGGCGCTTGGCATTTCTCGGTAAACGCAAGAAATTTCCCGGAGTT	1079	
Oy	1110	GGCAAAAGGTATAAAGATTCGCGGAGAGAGGTTGAGGTGATTTACGTCA	1169	
Db	1080	AGGAGAGCTTTACCGGATAAATTTGTTGCCGTTTATTTACGCTGGAAAT	1139	
Oy	1170	TAAAGGAAGTGAAGTTCTGTTTGAAGGGTCACTTCCAGAGAGAAATCCCG	1229	

3' UTR

Db 1140 TACTAATGGTAATCTCTGTATGACCGGAACTTTGATCCGGAGAAAGTAAGGGGAAGAT 1199
Qy 1230 GGTGATTTGATCGCGAGATCAATGGAAGATCGGAGAAAGGAGACCGGTTHAAGAAC 1289
Db 1200 TGTGATGTGATAGAGAAATTAATGCTAGAGTTTCAAGAAAGGTGATGTGGTTAAAGCAG 1259
Qy 1290 TGGAGGATTTGCAATGATCTAGCCATACAGAGATCAACACAGAGAGATCTATTGA 1349
Db 1260 AGTGGAGTTTGGAAATGATTTGGCTAATACCTGCGGGAATGGTGAAGAGCTTTTTCGGGA 1319
Qy 1350 GGTTCATCTCTTACAGCTACATGATTTGGTTACACTGAGTCACTGCTTCTGTAAGGCTTA 1409
Db 1320 TGCTCATTTGTTACCGGACCCCGCTTGTGTAAGACCGGTGATATATCGGCATTA 1379
Qy 1410 TGTTAATGCCACCGTGAACAAAGGCGCGGATTAATTTTGGTGGTACGGTGAATGGAG 1469
Db 1380 CGTGACTACTGATCTTAATCCACCGCTTCGATTTCAATCTTAGGAACAGCTCGTCGGTGT 1439
Qy 1470 GTACGAGCACCGAGGTGGTCACTAGTTTTCAGCTCGAGGACCGAGTTAGCCATCTTC 1529
Db 1440 TAAACCATCTCCGTTTTCGACGCGTTTAGCTACGTTGGACCGAATTCGATTAACCGAA 1499
Qy 1530 GATCTAAACCGGATGATTTGCTCGGAGTCAATATCATTCGCGCTTGGCTCAAAA 1589
Db 1500 TATTCCTTAAACCGGATCTGATCGCTCTGAGTAAACATCTCGCCGCGTGGACCGTGC 1559
Qy 1590 TCTAGGACCAACCGGACTTCTTATGATTCAGAGAGATTAATCTCACTGATATGTCAGG 1649
Db 1560 TGCTGGACCAACCGGACTCGTCTCGGATCTCTCGCGGTGGAGTTCAATATCATCTCTGG 1619
Qy 1650 AACTTCATGCTTGTGCACATGTTAGGGAATCACTGCTCTTATCCGGTCTCGATACCC 1709
Db 1620 GACGTCGATGCTTGTCCCTCAGTATAGTGTGTTAGCGCGGCTTCTCAAGTCGCTGCATCC 1679
Qy 1710 GAATGCTCTCAGCTGCAATCAATCCGATTCATGACACACAGCGGATTTAGCATCG 1769
Db 1680 TGAATGGAGCCGCGCGGATTAGATCGGCGCTTATGACACCGCTTACAAAACCTACAA 1739
Qy 1770 TCAAGGGAACCGATA-----AAGGATGGTAACAAACCGCGGTGTGTTCGATGG 1823
Db 1740 AGACGGTAAACGGTTACTCGACATCGCGACAGGAGAGCTTCGACGCGTTCGATCACGG 1799
Qy 1824 AGCAGGCGATGTGAATCGGAAAAGGCGATAAACCAGGATGGTTTACAACTTCAACC 1883
Db 1800 TGCAGGACACGTGTCAACCAACTGCCACTATCCAGGACTCATCTACCATCTAACGAC 1859
Qy 1884 AGTGATTAACATACTTACCTCTGCACTCTTGGATTACAGATCAGATATT 1935
Db 1860 GGAGGATTACTTAGCTTCTCTGCGATTGATTAACATACATCGCGCGAAATT 1911

RESULT 4
AF360285
LOCUS
DEFINITION
Arabidopsis thaliana 2618 bp mRNA linear PLN 18-SEP-2002
(At5g67360) mRNA, complete cds.
ACCESSION
AF360285
VERSION
AF360285.1 GI:3430745
KEYWORDS
FLI CDNA.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 2618)
Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Ban, J., Chung, M.K.,
Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Arabidopsis Full Length cDNA clones

TITLE

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 2618)
Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Ban, J., Chung, M.K.,
Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Direct Submission

TITLE
JOURNAL

Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
Arabidopsis Full-length cDNA': Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

COMMENT

The Salk, Stanford, PGECC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X.,
Sakano, H., Pham, P.K., Ban, J., Chung, M.K., Goldsmith, A.D.,
Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Chen, H.,
Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J.,
Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A.,
Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGECC)
contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to Genbank.

FEATURES
source

Location/Qualifiers
1..2618
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAFLOS-16-L13 (R09277)"
/note="This clone is in a modified pBluescript vector
(FLC-1) as a BamHI/XhoI insert.
ecotype: Columbia"

gene

1..2618
/gene="At5g67360"

5' UTR

1..114
/gene="At5g67360"

CDS

115..2388
/gene="At5g67360"
/note="contains subtilase domains PF00082 and a PA domain
PF02225"

/codon_start=1

/evidence=experimental

/product="putative subtilisin serine protease ARA12"

/protein_id="AAK25995.1"

/db_xref="GI:3430746"

/translation="MSSSLSTAFFLLCLGCFCHVSSSSDQGTIVVHAKSQMPSS
FDLHNSWYDSSLSISDAELLYYENAHGFTLQEEASLMTQPGVSVLPFHR
VELHTRPLFLGLDEHADIPEAGSYDVVGVLDTCVWPKESYSGDEGFPITPS
WKGCEAGTNTASLCKNLICARFARGVESTWGPIDESKESPRDDGHTHTSS
TAAGSVGASLLGASGTARGMAPRVAIVKVCGLGCPSSDLAIDKAIADNVN
VLSMSLGGSMDYRDGVAIGAFAMERGIIVSCAGNAGPSSSLSNVAPWITVGA
GLDLDFFALTLGNKQFTGVSLFKGALPKLLPFIYAGNATNGNLGCMGTLLI
PKVKGTVMACDRGINARVQKGVVKAAGVGMIILANTAAAGELVADAILPATV
EKAGDIIRHYVTDDNPNTASILGTVVGVKPSVWAAFSRGPNSITPNTILKPDLIA
PGVNILAWTGAAGTGLASDSRRVFNIIISGTSNCSCHVSLAALLKSVHPKSPAA
IRSLMTYATKYTKGKGLDIAATKPSFPDHGAGHVSPTATNPLGLIYDLTIEDYL
GFLCALNTSPQIRSVKRTTCDPSKYSVADLNYPFAVNVGVDGVAIKYTRIVTSV
CGAGTYSVKVTSETTGTGVRKIVSEPAVNFKEANEKKSYYTFTVDSKPSGNSFGSIE
WSDGKHVGVSPVAISWT"

2389..2618

/gene="At5g67360"

2602

3' UTR

/gene="At5g67360"

misc_difference

2602

BASE COUNT	/note="not present in genomic sequence"			
ORIGIN	542 a	632 c	628 g	716 t
Query Match	20.7%; Score 482; DB 8; Length 2618;			
Best Local Similarity	56.2%; Pred. No. 7.5e-126;			
Matches 974; Conservative	0;	Mismatches 740;	Indels 18;	Gaps 3;
Qy	210	ACTTCTCTACTCCTATGGCTCTCGGATTGAAGATTGTCGTCAGTTGACTGAATCAGA	269	
Db	306	ATTGCTCTACACTTACGAGAACGGATTTCATGGATTCTCAACACGGCTAACTCAAGAAGA	365	
Qy	270	AGCCGAGATCTAGATATTCACCTGAAGTTGTCAGTCAGACCTGACCATGTTCTTCA	329	
Db	366	AGCCGACTCGCTCATGACTCAACTCGTGTATATCTCCGTTTACCGGAGACCGTTACGA	425	
Qy	330	GGTTCAAACACACTTACTCTTACAAGTTCTTGGGACTCGACGGTTTTGGAACTCCGGTGT	389	
Db	426	GCTACACACACTCGTACT	482	
Qy	390	ATGGTCTAAATCTCGGTTTGGTCAAGCCAAATATTCGGCGTCTGTGATATCTGAGTTTG	449	
Db	483	GTTCCTTGAAGCGGCTCTTACAGCGACGTCGTGTCGGAGTTCTCGATACCGGAGTTTG	542	
Qy	450	GCTTGAAGTCTTAGCTTTGACGATACCGGAATGCCCTTCGATTCCACGGGAATCGAAGG	509	
Db	543	GCCCGAGATAAAGCTACTCCGACGAAGGATTTCGGTCTCTATTCCTTCTCTCGAAAGG	602	
Qy	510	GATTGGCAAGAGGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAAGCTAATCGGTGC	569	
Db	603	CGSATGCGAGCGCGGAACCAATTTTCAACGGCTTCTCTGTAAACGTAATATTCGGAGC	662	
Qy	570	TAGATTCTTCAATCAGAGGACACCGTGTGCTTAATTCACGAGGAAATCACCAAAATGCC	629	
Db	663	AAGATTCTTCTCGTGTGTAC-----GATCAACGATGGACCAATCATGATGATC	713	
Qy	630	TCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCGCCTCAACAGT	689	
Db	714	TAAAGATCAAGATCTCTTAGAGACGACGACGGTCAACGGAATCAACACCTCATCGACGC	773	
Qy	690	TGTTGGATCTCTGTTTCGATGGGAAATGTTCTTGCAATGAGCTGGTGTGCTCGTGG	749	
Db	774	CGCTGGATCCGCTGTTGAAGGAGCTAGCTTATTAGGCTACGTTTCAGGAACAGCTCGTGG	833	
Qy	750	GATGCTCTCGAGCTCACATTCAGTCTATAAAGTCTGTGTTCAATGTTGTTTACAG	809	
Db	834	TATGGCTCCACGCGCTCGTCTGCTGTTTACAAAGTCTGTGCTCGTGGTGTGTTTCAG	893	
Qy	810	CTCTGACATTTACGAGCTATAGATGTAGCGATTCAAGATAAAGTCGATGTTCTTTCGCT	869	
Db	894	CTCAGATATTTTACCTGCAATCGATAAAGCCATCGCGGATAACGTCATATGCTCTCGAT	953	
Qy	870	TTCCTTGGCGGTTTCCCTATTCCTTTGTATGATGACAAATCGCCATTGGAACTTCG	929	
Db	954	GTCACTCGCGGTGGAATGTCGGATTATTATAGAGACGGTGTTCGATCGGAGCAATCGC	1013	
Qy	930	AGCCATGAACCGGTATCTGTAATCTGTGAGCTGTTTACAAACGGTCCCAATCGAAG	989	
Db	1014	CGCCATGGAAGAGAGGATTTTAGTATCTTCTCAGCTGTAATGCTGGTCCGAGTAGTTC	1073	
Qy	990	CTCTGTTGCAACACAGCTCCTTGGGTCTCAACCATTTGGCGCAGGACGCTTGTATCGAAG	1049	
Db	1074	TAGTTTATCAACGTAGCTTCATGATACAACTGTTGCTGCTGTTAGATCTGTA	1133	
Qy	1050	ATTTCCCGCTGTGTCAGATTAGCCACCGGAAAGCTTCTCTATGGAGTCAATGATCC	1109	
Db	1134	TTTTCCGCGCTTGGATTCTCGGTACCGGAAGAAATTCACCGGAGTTTCTTTGTTTAA	1193	
Qy	1110	GGGAAAGGTATAAGATGCCGAGAGAGTTGAGTGATTTAGCTCACAGGAGAGA	1169	
Db	1194	AGGAGAAGCTTTACCGGATAAATTTGTTCCGCTTTTATTAGCTCGGAATGCTAGTAATGC	1253	
Qy	1170	TAAAGGAGTAGTCTCTGTTTGAGAGGTCATCTTCCAGAGAGAAATCCGAGGCAAAAT	1229	

1254	TACTAATGTAATCTCTGTATGACCGGAAC	111	TTC	CGGAGAAAGTAAAGGGGAAGAT	1313
1230	GGTGATTTGTGATCGCGGAGTCAATGGAAGAT	111	TTC	CGGAGAAAGGAGACCGGTTAAAGAAC	1289
1314	TGTGATGTGTATAGAGGAATTAATGCTAGAGT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1373
1290	TGGAGGAGTTTCAATGATCTTAGCCATA	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1374	AGTTGGAGTTTGAATGATTTCTGGCTAAC	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1350	CGTTTCATCTTTACCACTACATTTGATTT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1434	TGCTCATTTGTTACCGCGACACCGTTG	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1410	TGTTAATCCACCGGTGAACCAAGGCGCG	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1494	CGTGACTACTGATCTAATCCACCGCTT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1470	GTACAGACACCGGAGGTGGCTCAGTTT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1554	TAAACATCTCGGTTGTGCGAGGTTTAG	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1530	GATACATAAACCAGGATATGATTTGCT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1614	TATTTCTTAAACCGGATCTGATCGCT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1590	TCTAGGACCAACCGGACTTCTTATGAT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1674	TGCTGGACCAACCGGACTTCTTCCGAT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1650	AACTTCAATGCTTGTCCACATGTTAG	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1734	GACGTGATGTTTTCGCTCAGTTAGT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1710	GAACTGGTCTCCAGCTCAATCAAAAT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1794	TGAATGGAGCCCGCGCGGATTAGT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1770	TCAAGGGAAGCGATA-----AAGGAT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1854	AGACGGTAAACGGTTTACTCGACAT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1824	AGCAGGCAATGGAATCCGAAAAGG	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1914	TGCAGGACACGTGTCCACCAACACT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1884	AGTGGATTACATACTTACCTTCGACT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
1974	GGAGGATTACTTAGCTTCTCTGCGAT	111	TTC	CGGAGAAAGGAGTGTGTTAAAGCAGC	1433
RESULT 5	AF055639	2686 bp	DNA	linear	PLN 14-AUG-2000
LOCUS	Arabidopsis thaliana cucumisin-like				
DEFINITION	complete cds.				
ACCESSION	AF055639				
VERSION	AF055639.1	GI:3176873			
KEYWORDS	Arabidopsis thaliana (thale cress)				
SOURCE	Arabidopsis thaliana				
ORGANISM	Burkhardtia; Virdiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.				
REFERENCE	1 (bases 1 to 2686)				
AUTHORS	Ribeiro,A., Akkermans,A.D., van Kammen,A., Bieseling,T. and Pawlowski,K.				
TITLE	A nucleotide-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development				
JOURNAL	Plant Cell 7 (6), 785-794 (1995)				
MEDLINE	95375542				
PUBMED	7647567				

Db	1254	TACTAATGGTAAATCTCTGTATGACCGGAACCTTTGATCCCGAGAAAGTAAAGGGAGAT	1313		
Qy	1230	GGTGATTTGTGATCGCGGAGTCAATGGAAGATCGAGAAAGGAGAGCGGTTAAAGAGC	1289		
Db	1314	TGTGATGTTGTATAGAGAAATTAATGTAGAGTTCAAGAAAGTGTGTTAAAGCAGC	1373		
Qy	1290	TGGAGGAGTTTCAATGATCTTAGCCAATACAGAGATCAACCAAGAAAGATTTATTTGA	1349		
Db	1374	AGTGGAGTTGGATGATCTTGGCTAACACTCGCGCGAATGGTGAAGAGCTTTGTCGGA	1433		
Qy	1350	CGTTCACTCTTTACCACTACATTTGATTTGTTACATGAGTCAGTCTCTCTGAAAGGCTTA	1409		
Db	1434	TGCTCATTTGTTACCGGCGACCGGTTGTGAAAAAGCCGGTGATATAATCCGSCATTA	1493		
Qy	1410	TGTTAATGCCACGGTGAAACCAAGAGCGCGATTAATTTTGGTGTACGGTGTGAGG	1469		
Db	1494	CGTGACTACTGATCTTAATCCACCGCTTCGATTTCAATCTTAGGAACAGTCGTCGTTGT	1553		
Qy	1470	GTCAGGACCGGAGGTGGCTCAGTTTTCAGCTCGAGGACCGAGTTTAGCCAATCCTTC	1529		
Db	1554	TAAACCATCTCCGTTTTCGAGCGTTTAGCTCAGTGGACCGAATTCGATTACACCGAA	1613		
Qy	1530	GATACATAAACCGGATATGATTGCTCCGGAGTCAATATCAATTCGGGCTTGGCCTCAAAA	1589		
Db	1614	TATTTCTTAAACCGGATCTGATCGCTCTCGAGTAAACATCTCTCGCGGTGGACCGTGC	1673		
Qy	1590	TCTAGACCAACCGGACTTCTCTATGATTCAAGAGAGTTAACTTCACTGTAATGTGAG	1649		
Db	1674	TGCTGGACCAACCGGACTCGCTTCGGATTCTCGCGCGGTGGAGTTCAATATCATCTCTGG	1733		
Qy	1650	AACCTCAATGTTGTTCCACATGTTAGCGGAATCACTGCTCTTATCCGGTCTGCATACC	1709		
Db	1734	GAGCTCGATGTTGTCCTCCTCAGTTAGTGGTTTAGCGCGCTTCTCAAGTCGGTGCATCC	1793		
Qy	1710	GAACTGGTCTCCAGTGCATCAATCAATCCGATTCATGACAAACAGCGGATTTGTACGATCG	1769		
Db	1794	TGAATGGAGCGCGCGCGGATTAGATCGCGCTTATGACCAACCGCTTACAAAACCTACAA	1853		
Qy	1770	TCAAGGAAAAGCGATA-----AAGGATGTTAAACAAACCGCGGTGTTGGATTTGG	1823		
Db	1854	AGACGCTTAAACCGTTTACTCGACATCGCGACAGGAAAGCTTCGACCGCTTCGATCAGG	1913		
Qy	1824	ACGAGGATGTAATTCGCAAAAGCGGATTAACCCGGATTTGTTTACACATTCACAC	1883		
Db	1914	TGCAGGACACGTTGTCACCAACAACTGCCACTTAATCCAGGACTCATCTACGATCAACGAC	1973		
Qy	1884	AGTGGATTACATAACTTACCTCTGCACTCTTGGATTCAAGATCAGATATT	1935		
Db	1974	GGAGATTACTTAGGCTTCCTCTCGCATTTGAATTACACATCGCGCAAAAT	2025		

RESULT 5	AP065639	2686 bp	DNA	linear	PLN 14-AUG-2000
LOCUS	Arabidopsis thaliana cucumis-like serine protease (ARA12) gene,				
DEFINITION	complete cds.				
ACCESSION	AF055639				
VERSION	AF055639.1	GI:3176873			
KEYWORDS	Arabidopsis thaliana (thale cress)				
SOURCE	Arabidopsis thaliana				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 2686)				
AUTHORS	Ribeiro,A., Akkermans,A.D., van Kammen,A., Bisseling,T. and Pawlowski,K.				
TITLE	A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development				
JOURNAL	Plant Cell 7 (6), 785-794 (1995)				
MEDLINE	95375542				
PUBMED	7647567				

510 GATTTCGCAAGAGGAGAAGTTTCAGTCTCTTCAGCTGTAAACCGGAAGCTAATTCGGTGC 569

778 CGGATGGAGCGCGGAACCAATTCACCGCTCTCTCTGTAAACGCTAAACTAATTCGGAGC 837

570 TAGATTCTTCATCAGAGGACACCGTGTGCTGAATTCACAGAGGAATCCAAACATGCC 629

838 AAGATTCTTCGCTCGTGGTTAC-----GAAACAAGATGGGACCAATTCGATGAATC 888

630 TCGTGAATACATTTCCGCAAGAGATTCACGGGACACGGGACTCACACGGCTCAACAGT 689

889 TAAAGAAATCAAGATCTCCTAGAGACGACGACGGTCAACGAACTCACACTCATCGACGC 948

690 TGGTGGATCTCTGTTTTCGATGCGCAATCTTCTGGCAATGGAGCTGCTGTGGTTCGTTG 749

949 CGCTGGATCCGTCGTTGAAGGAGCTAGCTTATTAGCTACGCTTCAGGAAAGCTCGTGG 1008

750 GATGCTCCTGGAGCTCACATTGCACTCTATAAGTCTGTGGTTCAATGTTGTTCACAG 809

1009 TATGCTCCACGCGCTGCTGCTGTTTACAAAGTCTGTTCGGCTCGGTGGTGTGTTTCAG 1068

810 CTCTGCAATTCCTAGCAGCTATAGATGTAGCGAATCAAGATAAAGTCGATGTTCTTCGCT 869

1069 CTCAGATATTTTAGTTCGATCAATCAATAAGCCATCGCCGATAACGCTCAATGCTCTGTCGAT 1128

870 TTCCCTTGGCGGTTCCCTATTCTCTTGATGATGACACAATCGGCATTCGGAACATTCGG 929

1129 GTCACTCGCGCGTGGAAATGTCGGATATTATTAGAGCGGTGTGGATTCGGAGCAATTCGC 1188

930 AGCATGGAAACGCGGTATATCTGTAATCTGTGCAGCTGGTAAACAACGGTCCAATCGAAAG 989

1189 CGCCATGGAAGAGGGAATTTAGTATCTTCCTCAGCTGGTAAATGCTGTGTCGGAGTAGTTC 1248

990 CTCTGTTGCAACACAGCTCCTTTGGGTCTCAACCATTTGGCGGAGGACGCTTGATCGAAG 1049

1249 TAGTTTATCAAAACGTAGCTCATGGATCAAACTGTTGCTGTGCTGCTACTTTAGATCGTGA 1308

1050 ATTTCCCGCTGTGTCAGATTAGCCAAACGGAAGCTTCTCTATGGAGAGCTCAATTCGTATCC 1108

1309 TTTTCCGCGCTTCGGAATCTCGGTAAACGCAAGAAATTCACCGGAGTTTCTTTGTTAA 1366

1110 GGGAAAGGTATAAGAAATCGCGGAGAGAGGTTGAGGTGATTACGTCAACAGGAGGAGA 1166

1369 AGGAGAACTTTACCGGATAAATGTTGCGCTTTATTACGTGGGAATGCTAGTAATGC 1426

1170 TAAAGGAAGTCAGTTCCTGTTTGAGAGGCTCACTTCCAGAGAGAATAATCCGAGGCAAAAT 1226

1429 TACTAAATGTAATCTCTGTATACCGGAACTTTTGATCCCGGAGAAAGTAAGAGGGAAGAT 1488

1230 GGTGATTTGTGATCGCGGAGTCAATGGAAGATCGGAGAAAGGAGAGCGGTTAAAGAGAC 1288

1489 TGTGATGCTGATAGAGGAATTAATGCTAGAGTTCAGAAAGGTGATGTCGTTTAAAGCAGC 1544

1290 TGGAGGAGTTCGAATGATCTTTAGCCAATACAGAGATCAACCAAGAAGAAGATTCATTGA 1344

1549 AGGTGGAGTTGGAATGATCTGGCTAACACTGCGGCGAATGCTGTAAGAGCTGTTGCGGA 1608

1350 CGTTCATCTCTTACCAGTCAATTAAGTTGTTACATGATCAGTCCCTTCTGAGGCTTTA 1408

1609 TGCTCATTTGTATACGGGACCAACCGTTGTTGTAAGAAAGCGGTGATATAATCCGCAATTA 1666

1410 TGTTAATGCCAGGTGAAACCAAGGCGGGAATAATTTTGGTGGTACGCTGATTCGGAG 1466

1669 CGTGACTACTGATCTCAATCCGACCGCTTCGATTTCAATCTTAGGAACAGTCCGTCGGTGT 1726

1470 GTCAAGAGCACGGAGGTGGCTCAGTTTTTTCAGCTCGAGGACCGAGTTTAGCCAAATCCTTC 1526

1729 TAAACCATCTCGGTTGCGAGCGTTTAGCTCAGCTGAGGCCGAATTCGNATTACCGGA 1786

1530 GATACTTAAACCGGATATGATTGTTCCGGAGTCAATATCAATTCGGCTTGGCTCAAAA 1586

1789 TATTCTTAAACCGGATCTGATCGCTCCTCGGAGTAAACATCTCTCGCGCGTGGACCGGTGC 1846

1590 TCTAGGACCAACCGGACTTCCTTATGATTCGAAGAGAGTTAACTTCACTCTAATGTCAGG 1646

2 (bases 1 to 2686)
 Yamagata,H., Uesugi,M., Saka,K., Iwasaki,T. and Aizono,Y.
 Molecular cloning and characterization of a cDNA and a gene for
 subtilisin-like serine proteases from rice (*Oryza sativa* L.) and
 Arabidopsis thaliana
 Biosci. Biotechnol. Biochem. 64 (9), 1947-1957 (2000)
 20506592
 11055401
 3 (bases 1 to 2686)
 Yamagata,H.
 Direct Submission
 Submitted (13-MAY-1998) Laboratory of Biochemistry, Faculty of
 Agriculture, Kobe University, Nada, Kobe 657-8501, Japan
 Location/Qualifiers
 1..2686
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 <290..>2563
 /gene="ARA12"
 <290..>2563
 /gene="ARA12"
 /product="cucumis-like serine protease"
 290..2563
 /gene="ARA12"
 /note="similar to subtilisin-like protein encoded by
 GenBank Accession Number X85974; intronless gene"
 /codon_start=1
 /product="cucumis-like serine protease"
 /protein_id="AAC18951.1"
 /db_xref="GI:3176874"
 /translation="MSSSFLSTAFPLLLCLGFCVHSSSSDQCTIYVHMKSQMPSS
 FPLHNWSDSLRSDSDELLTYENAIGHFSTRLTQEEADSLMTQGEIVSLVEPHR
 YSLHTRTFLFLGDHETFLPEAGSYGDVVGVLDITGWPEKSPQSYDEGGPIPS
 WKGCCAGNFTASLGNRKLIGARFARGYVETMGPIDESKESRPRADDGGHTSTS
 TAGSVVEGASLLGYASGTARGMAPARVAVYKVLGCGFSSDILAAIDKAIADNVN
 VLSMUGGWSDIYRQVAGLGAFAAMERGLVSCAGNAGPSSSLSNVAPLITVGA
 GTLDFRPAALINGKNFTVSLFKGEALPKLLPFIYAGNANLNCNLCMTGLI
 PEKVKGIVMCDRGINARVQKDVKAAGVGIMLANTAANGEELVADHLLPATTVG
 EKGADIIHYVTTDPNPTASISLTGVGWGSPVVAAPSSRGPASTPNILKPLLIA
 PGVNIILAAATGAGPTGLASDSRRVEFNIISGTSMSCPHVSGLAALLKSVHPWSPAA
 IRSALMTAYKYKDKPLLDIATGKSTFPFHGAGHVSPTTATNGLIYDLITDYL
 GFCLANYSPQIRSVRRNYTCDPSKYSVADLNPFPFVAVNDVGAGVAKYTRTWSV
 GGATYKSVKTTSETTKVTSVEPVLNFKEANKEKSYTFTFTVDSKPSGSSNSFGSIE
 WSDGKHVGVSPAISMT"
 BASE COUNT 687 a 646 c 639 g 714 t
 ORIGIN
 Query Match 20.7%; Score 482; DB 8; Length 2686;
 Best Local Similarity 56.3%; Pred. No. 7.5e-126;
 Matches 974; Conservative 0; Mismatches 740; Indels 18; Gaps 3;
 QY 210 ACTTCTCTACTCTATGGCTCTCGGATTGAAGATTGCTGCTCAGTGTCACTGAATCAGA 269
 Db 481 ATTGCTCTACATTACGAGACCGATTCTATGATTCTCAACACGGCTACTCAAGAAGA 540
 QY 270 AGCGGAGACTAGATATTCACTGAAGTGTGTGAGTCAGACCTGACCATGTTCTTCA 329
 Db 541 AGCCGACTCCTCATGACTCAACTGGTGTATCTCCGTTTACCGGAGCACCGTTACGA 600
 QY 330 GGTTCRAACCACTTACTCTTACAAGTTCCTTGGGACTCGACGGTTTTTGGAAACTCCCGTGT 389
 Db 601 GCTACACCACTCTGACTCTCTCTCTCTCGGTCTCAGC---AACACACCGAGATCT 657
 QY 390 ATGGTCTAAATCTCGGTTTGGTCAAGGCACAAATTATCGCGCTGCTTGATCTCGAGTTTG 449
 Db 658 GTTCCCTGAAGCGGCTCTTACACGACGCTCGTCTCGGAGTTCTCGATACCGGAGTTTG 717
 QY 450 GCCTGAAGTCTTAGCTTTTGACGATACCGGAATGCCTTCGATTCACCGGAATGGAAGG 509
 Db 718 GCCCGAGTAAAGACTACTCCGACGAAGAATTCGGTCTTATTCCTTCCTCTCGGAAGG 777

```

1849 TGCTGACCAACCGGAGCTCTCGGATTCCTCGCGCGTGAGTTCATATCATCTCTGG 1908
1650 AACTTCAATGCTGTGCACATGTTAGCGGAATCACTGCTCTTATCGGCTCGATACCC 1709
1909 GACGTCGATGCTTTCCTCCCTACGTTAGTGGTTAGCGCGGCTTCTCAAGTCGGTGATCC 1968
1710 GAACCTGGTCTCCAGCTGCAATCAATCCGCAATGATGACACACGCGGATTTGACGATCG 1769
1969 TGAATGGAGCCCGCGGAGTATGATCGCGCTTATGACACCGCTTACAAACCTACAA 2028
1770 TCAAGGGAAGGATA-----AAGGATGGTAACAAACCGCGCTGTGTTGCGATGG 1823
2029 AGACGCTAAACCGTTACTCGACATCGACACGAGGAGCCTTCGACGCGGTTGATCACGG 2088
1824 AGCAGGCGATGTAATCGCAAAACCGGATTAACCGCGGATTTGTTACACATTCACAC 1883
2089 TCGAGGACAGCTGTCAACCAACTGCCACTATCCAGGATCTATCAGATCTAAGCAC 2148
1884 AGTGGATTACATAACTTACCTCTGCACTCTTGGATTCAACAGATCAGATATT 1935
2149 GGAGGATTACTTAGGCTTCTCTGCGCATGGAATTACACATCGCGCAAAATT 2200

```

RESULT 6

```

AB007645/c
LOCUS
DEFINITION
AB007645 Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:k8K14.
ACCESSION
AB007645.1 GI:2564045
VERSION
AB007645.1
KEYWORDS
SOURCE
ORGANISM

```

Arabidopsis thaliana (thale cress)

```

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (sites)
Nakamura,Y., Sato,S., Kaneko,T., Kotani,H., Asamizu,E., Miyajima,N.
and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 5. III.
Sequence features of the regions of 1,191,918 bp covered by
seventeen physically assigned P1 clones
DNA Res. 4 (6), 401-414 (1997)
98162728
9501997
2 (bases 1 to 72698)
Nakamura,Y.
Direct Submission
Submitted (03-OCT-1997) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agg_graph.cgi?c=k8K14
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://grenlil.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Rddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K3G17 and the 3' clone is K919.
Location/Qualifiers

```

REFERENCE

```

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

```

COMMENT

```

For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agg_graph.cgi?c=k8K14
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://grenlil.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Rddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K3G17 and the 3' clone is K919.
Location/Qualifiers

```

source

```

1. 72698
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="K8K14"
/clone_lib="Mitsui TAC"
complement(join(283..653,770..1619))
note="contains similarity to unknown protein"
gene_id:k8K14.1
sp|Q10058
/codon_start=1
/evidence=not experimental
/protein_id="BAB09014.1"
/db_xref="GI:9758428"
/translation="MAVISLLPFPVLTTRITICNSSKSMAAKQIGSETSKRVFV
CGGVIQVGTAYFLAKGIATLVKSAACASAGGAGFLAFDWCGGDFVAFHARAS
FKHSLAEELNGVSYGYRALTTLSVTYTESKPGGGGLPDWNGPAAKSTIGTIT
QTAAVHPQLPRLKLSITAREKYGVEVYKLEEVESGVRVNSVLTVEGVRVITDVT
VLAMPKSKFELLSSIFRYCTKAHSIVLEKPEPAITPHALFLYRAGGEALDP
VAMPPTGEVYICGMSQOEVPDDAQVTSNPESIOVLKRVKTVSSVYNEALVKA
EQACFLPSTEDGVPVIGIEPGIKGCVGTGHCWGILNGPATGAALAEILVDVATSV
DLRSPSPSRFSKRR"
5973..6890
/note="gene_id:k8K14.2
pir||S71284"
/codon_start=1
/evidence=not experimental
/product="myb-related protein, 33.3K"
/protein_id="BAB09015.1"
/db_xref="GI:9758429"
/translation="MADRIKGPSPDEOLRELVLVYKGRNWTVISKIPGRSGKSC
RLRNCNLSQVDEHPESAEDETIARAHAQCNKATARLLNGTIDNAKNHWNST
LKRKCGYDHRGVDGSEDEHPKRSYAGSPVVTGLYMSPTGSDYSDSTIPL
PSVLPFRPFRGAVLPLPFISSSDSDPTSLSLPGAQVSESNKSHSSTINN
TTSRHHNTVTFMFPFSGGFRGAIEMGKSPFGNGGFMAVYQEMIKAEVSYMTM
QRNNGGFGVGFIDNGMIPMSQICVGRIE"
complement(join(7177..7782,7871..8755))
/note="gene_id:k8K14.3"
/codon_start=1
/evidence=not experimental
/product="cytochrome P450"
/protein_id="BAB09016.1"
/db_xref="GI:9758430"
/translation="MGEMIDLFLLMAGFIAVAVFRSKOKNLPNPVPRVPIGHL
HLKEPVHSLRDLNLGIDVILRIGSERAVVTASAAEFLSQQDVVFAKPL
ATLEYMGXNLTIVSTAPYGEHRRRLRPFCAVDILSTARLDFDRDEVRAMIRKI
NVELVTGGSVRLKLOPFLYLTYNILMSVAGKREDEETKEVRKLREVFAGVN
YVGFPLTLKLDLGYRKRAKLAGLDFKQKLVDEHKKRGAELKFTWITRLLS
LQSEPECYTDDIIGLVQVMLAGTDTTAVTLEWAMNLLNHPVLRKLTLENS
KEGVRPESDTGKCPYLNNTVSETLRFPAAPLLVPHASSTCEVAGFDIPLRTWLF
NAAIQRDPNVDDPETFKEPSESETHRGKFLPFGIGRACPGMLAQVLVSLGVS
LIQCFMERNDVAVMSGGLTMRKSVPLVAKCSLFLDKVLV"
join(9132..9192,9618..9763,9851..10349,10429..10523,
10630..10676,10761..10911,11066..11129,11503..11661,
11749..11870,11983..12057,12141..12194,12290..12363,
12464..12538,12619..12838)
/note="gene_id:k8K14.4
ref|NP_005638.1
strong similarity to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB09017.1"
/db_xref="GI:9758431"
/translation="MSLSVSELVNLFVRLQBSGTHAATFLGYEAGINKSNIDGNM
VPGALLKFKVQKGLQYMEANLSEVDIDFSPFDLLISKDVRELQDMLREKK
KRRDKMEKDRSKENDKGVREHEDGRNRAKEKREKEREKEREKEREKEREK
EKIEREKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREK
EMVTDQTKRIAGDGTGAPMDIVMTPTSTQSHIPNSDVRILEGHSVCAWAPS
ASLSSAGSGATARIWSIPGSPKAVHTGTNNALILKAKGSKNEKSKDVTLDWNG
EGTLLATGSCDGGARITWTLNGELISTLSKHPGFISLKNKXGDIYLLTGSVRTAVW
DVRAEWKQQFHFSGFTLDDVNRNNSVSTATSDSMIYLCKIGETRPARTFTGHOGE

```

CDS

CDS

CDS

CDS

FEATURES

Db 25250 AGGAGAGCTTTACCGGATAAATGTTGCGGTTATTTACGCTGGGAATGCTAGTAATGC 25191
Qy 1170 TAAAGGAAGTGAGTCTGTTTGGAGAGGTCACCTTCAAGAGAGAAATCCGAGGCAAAAT 1229
Db 25190 TACTAATGTAATCTCTGATGACCGGAACTTTGATCCCGGAGAAAGTAAGGGAGAT 25131
Qy 1230 GGTGATTTGATCGGGAGTCAATGAAGATCGGAGAAAGGAGAAAGCGGTTAAAGAAC 1289
Db 25130 TGTGATGTTGATAGAGAAATTAATGTAAGTTTCAAGAAAGTGTGTTTAAAGCAGC 25071
Qy 1290 TGGAGGATTTGCAATGATCTTACCAATACAGAGATCAACCAAGAGAAAGATCTTATGCA 1349
Db 25070 AGGTGGAGTTTGGAAATGATCTTGGCTAACACTCGGGCGAAATGGTGAAGAGCTTGTTCGCGA 25011
Qy 1350 CGTTCATCTCTTACAGCTACATTTGATGTTTACACTGAGTCACTGCTTCTCAAGGCTTA 1409
Db 25010 TGCTCATTTGTTACCGCGACACCGCTTGTGTAAGACCGGATATTAATCCGGCAATTA 24951
Qy 1410 TGTTAATGCCACGGTGAACCAAGCGCGGATTAATTTTGGTGGTACCGGATTTGGGAG 1469
Db 24950 CGTGACTACTGATCCTAATCCACCGCTTCGATTTCAATCTTAGGAACAGTCTCGGTGT 24891
Qy 1470 GTACGAGACCGGAGTGGTCTAGTTTTCAGCTCGAGGACCGAGTTTAGCCNAATCCTTC 1529
Db 24890 TAAACCATCTCCGGTTTCTCGACGGTTTAGCTCACGTGGACCGAAATTCGATTAACCGAA 24831
Qy 1530 GATACATAAACCGGATATGCTCCGGGAGTCAATATCATTTCCGGCTTGGGCTCAAAA 1589
Db 24830 TATTCCTTAAACCGGATCTGATCTGCTCTGAGTAACATCTCGCGCGTGGACCGGTGC 24771
Qy 1590 TCTAGGACCAACCGGACTTCTTATGATTTCAAGAAAGATTAACCTCACTGTAATGTGAG 1649
Db 24770 TGCTGGACCAACCGGACTCGCTTCCGATTTCTCGCGCGTGGAGTTCAATATCATCTCG 24711
Qy 1650 AACTTCATGCTTTGTCACATGTTAGGGAATCACTGCTCTTATCCGGTCTGCAATCC 1709
Db 24710 GACGTCGATGCTTTGCCCTCACGTTAGTGGTTTAGCGCGCTTCTCAAGTCGGTGCATCC 24651
Qy 1710 GAACGTGCTCCAGCTGCAATCAATCCGATTTGATGACACACGCGGATTTGACCATCG 1769
Db 24650 TGAATGGAGCGCGCGGATTTAGATCGGCGCTTATGACACCGCTTACAAAACCTACAA 24591
Qy 1770 TCAAGGAAAGCGATA-----AAGGATGTTAAACAAACCGCGGTGTGTTGCGATGG 1823
Db 24590 AGACGTTAACGTTTACTCGACATCGCGACAGGAGAGCTTTCGACGCGTTCGATCACGG 24531
Qy 1824 AGCAGGCGATGTAATCCGAAAAGGCGAATAAACCAGGATTTGTTTACAACTTCAACC 1883
Db 24530 TGCAGGACACGTTGTCACCAACACTGCCACTAATCCAGGACTATCTACGATCTAACGAC 24471
Qy 1884 AGTGATTAATTAATCTTACCTCTGCACTCTTGGATTTCAAGATCAGATATT 1935
Db 24470 GGAGGATTAATCTTACGCTTCTGCGATTAATTAACATCGCGCGCAATTT 24419

RESULT 7
AX506370 2265 bp DNA linear PAT 27-SEP-2002
LOCUS Sequence 1065 from Patent WO0216655.
DEFINITION AX506370
ACCESSION AX506370
VERSION AX506370.1 GI:23387607
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
TITLE Stress-regulated genes of plants, transgenic plants containing same, and methods of use
JOURNAL Patent: WO 0216655-A 1065 28-FEB-2002;

The Scripps Research Institute (US) ; Syngenta Participations AG
(CH)
FEATURES
source Location/Qualifiers
1..2265 /organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 519 a -510 c 542 g 694 t
ORIGIN
Query Match 20.7%; Score 481.8; DB 6; Length 2265;
Best Local Similarity 55.3%; Pred. No. 8.5e-126;
Matches 1033; Conservative 0; Mismatches 812; Indels 24; Gaps 4;
Qy 380 ACTCGGTGATGCTTAATCTCGTTCGTCAGGCAAAATATATCGGCGTGTGATA 439
Db 344 AATTCGGTGTTCACGATCTCGGTTCTTCTTCAACGGCGTTATCATCGGAGTTTAGATA 403
Qy 440 CTGGAGTTTGGCCGTAAGATCTAGCTTTCAGATACCGGAATGCCCTTCGATTCACCGGA 499
Db 404 CTGGGTTTGGCCCTGAACTTAGAAGCTTCGATGATCTATATGCTGAGATTCCTTCTA 463
Qy 500 AATGGAAGGATTTGCCAAGAGGAGAAAGTTTCAGTCTTTCGAGCTGTAAACCGGAAGC 559
Db 464 AATGGAAGGAGAAATGGAATCTGTTCCGATTTTCGATTCGAAGTTTGTAAACAAGAAC 523
Qy 560 TAATCGGTGCTAGATTTCTCATACAGAGACACCGTGTGCTAATTCACAGAGAAATCAC 619
Db 524 TTATCGGAGCTAGAAGCTTCTCCAAAGGATTTCAATTCGGCTTCTGGTGGTGTTCGA 583
Qy 620 CAACATCCCTCGTGAATACATTTCCGCAAGAGATTCACCGGACACGGGACTCACACCG 679
Db 584 GTAA-----CGGTGAATCTGTTCTCTCTGATGTTGACGGACATGGAACATACTT 637
Qy 680 CCTCAACAGTTTGGTGGATCTCTGTTTCGATGGGGAATGTTCTTGGCAATCGAGCTGGTG 739
Db 638 CAACTACCGCGCGGATCCCGCTTAGAAGCTAGCTTCTCGTTTACGCCCGGTA 697
Qy 740 TGGCTCGTGGAGTCTCCTGGAGCTACATTCAGTCTATAAAGTCTGTGTTGTTCAATG 799
Db 698 CGGCGAGAGGTATGSCCACTCGTCTGTTGTTGTTTACTTATAAAGTTTGTGGAGTACTG 757
Qy 800 GTTGTTCAGCTTCGACATTTAGCAGCTATAGATGATGCGATTCAGATTAAGTCGATG 859
Db 758 GTTGTTTGGATCTGATATCTAGCTGTATGGATCGAGCTATATCTGATGGTGTGATG 817
Qy 860 TTCTTTTCGTTTCCCTTCGGGTTTCCCTATTCTTCTTATGATGACACAATCGCATTG 919
Db 818 TGCTTTCTGTTATCTCTTGGTGGTGTCTGCTCCGATTTATCGGATACGATTCGATG 877
Qy 920 GAACATTCGAGCCATGGAACGCGGTATATCTGTAACTGTGTCAGCTGTTAAACCGTTC 979
Db 878 GAGCGTTTTCGGCTATGGAGAGAGTGTGTTTGTGCTTCTGCTGTTAATAGTGTGTC 937
Qy 980 CAATCGAAAGCTCTGTTGCAACACAGCTCTTGGGCTCTCAACATTTGGCGGACGACGC 1039
Db 938 CTACTAGAGCTTCTGTTGCCAATGTTGCTCTTGGGTTATGACTGTTGGTGTGCTACTT 997
Qy 1040 TTGATCGAGATTTCCGCTGTTGTCAGATAGCCAAACGGAAGCTTCTCTATGAGAGT 1099
Db 998 TAGATAGAGATTTCCGGCTTTTGGCAATCTCGGTAACGGGAACGACTTACCGGTGTTT 1057
Qy 1100 CATTGTATCCGGGAAAGGTATAAAGAAATCCGGGAGAGAGGTTGAGGTGATTTACGTCA 1159
Db 1058 CGCTGTATAGCGGTGTAGAAATGGGGA-----CGAAGCGCTTGAATTTGTTTATAATA 1111
Qy 1160 CAGGAGGAGATAAAGAAAGTGTGCTGTTTGAAGGGTCACTTCCAGAGAAAGAAATCC 1219
Db 1112 AAGGGAATAGTAGTTCGAGTAATCTTGTTTACCTCGTTCGCTTGTATTCAGTATTGTTTC 1171
Qy 1220 GAGGCAAAATGTTGATTTGTGATCCGGAGTCAATGGAAGATCGGGAAGAGAGAGCGG 1279
Db 1172 GTGGGAAGATTTGTTGTTTGTGATAGAGGTGTTTAATGCTAGAGTTGAGAAAGGAGCTGTGG 1231

Qy	1280	TTAAAGAACTGGAGAGTTGCAATGATCTTTAGCCAATACAGAGATCAACCAAGAAG	1339
Db	1232	TTAGAGATCTGTTGGTTAGGATGATAATGGCGAATACTGCTGCGAGTGGAGAGC	1291
Qy	1340	ATTCTATTGACGTTGATCTCTTACAGCTACATTTGATTGGTTACACTGAGTCAGTCTTC	1399
Db	1292	TTGTGCGGATAGTCAATTTGCTTCCCGGATCGCTGTAGGGAAGAAGACTGGTATTAC	1351
Qy	1400	TGAAGCTTATGTTAATGCCAGCGTGAACCAAGGCGCGGATAATTTTTGGTGGTACGG	1459
Db	1352	TTAGGAGATGTTAATGTCAGATTCTAAACCAACCGCTCTCTTCTTTTAAAGGAACGG	1411
Qy	1460	TGATTTGGGAGTCAAGACACCGGAGTGGCTCAGTTTTCAGCTCGAGGACCGAGTTTAC	1519
Db	1412	TTCTTCACGTTAAGCGCTCTCTGTTGCTGTTTGTCTTTAGCTCGAGAGTCCCTAATAC	1471
Qy	1520	CGAATCTTCGATACATAAACCGGATATGATGCTCCGGAGTCAATATCATTTGCGGCTT	1579
Db	1472	TTACTCTCGAATCTTTGAACCGCTGATGTTATTTGGTCTCGAGTTAATATTTGGCTGTT	1531
Qy	1580	GGCTCAAAATCTAGGACCAACCGGACTTCCTTATGATCAAGAAGAGTTAACTTCACTG	1639
Db	1532	GGTCTGACGCTATGCTCTCTGTTGCTGTTGACAGGACTCTAGAGAGACTCAGTTCAACA	1591
Qy	1640	TAATGTCAGGAATCTCAATGCTCTTGTCCACATGTTTAGCGGAATCAGTCTCTTATCCGGT	1699
Db	1592	TCATGTCAGTAGCTCAATGTCATGCCACACATCAGTGGTTTAGCGGCTCTTTTGAAG	1651
Qy	1700	CTGCATATCCGAAGTCTGCTCAGCTGCAATCAATCCGATTCGATGACACACCGGATT	1759
Db	1652	CAGTCAACCTGAGTGGAGTCCGAGTGTCTCAATCAGCTCTCATGACTACAGCTTACG	1711
Qy	1760	TGTACCATCGTCAAGGGAAGGATAAAGGATGGT-----AACAAACCCAGCGGTGTGT	1813
Db	1712	TTCTTGACAAACCAACCGCTCTCTCCATGATGCTGCGAGCAACAGCGCTATCTTAACCAT	1771
Qy	1814	TTGCGATTGAGCAGGGCATGTGAATCCGCAAAAGCGGATAAACCCGGGATTGTTTACA	1873
Db	1772	ATGCTCAGGCTCGGGCCATGTAGATCCCAAAAGGCTCTCCACGAGTCTTGTCTACG	1831
Qy	1874	ACATCAACAGTGGATTCATACTTACCTCTGCACTCTTGGATTTCAGAGATCAGATA	1933
Db	1832	ACATCTCAACCCGAGGAATACATCAGGTTTGTGCTCTAGACTACACAGTCGATCACA	1891
Qy	1934	TTTATGCAATCACTATAGACGTCAGCTGCAATGGATATTCGGGAATCCCGGTT	1993
Db	1892	TTGTTGCGAT-----TGTGAACGACCTAGCGTTAACTGCTCGAAGAGTCTTCAGATC	1945
Qy	1994	TTAGTCTCAATTACCCGTCGATGCGGTGATTTTCAACCGTGGCAAGACTACGGAGATGA	2053
Db	1946	CTGCTGAGCTCACTACCAAGTTCTCGTTTGTGTTTGTGGGGTAAAGAGTTGTGCGGT	2005
Qy	2054	TCACAGGCGTGTCTATAAGTGGGAGTCTTAATCTCGATATACATGATGTAATGTCAAGG	2113
Db	2006	ACACTCGGAAGTAACAAATGTTGGTGCAGCAAGCTCGGTTTACAAAGTACGGTTAATG	2065
Qy	2114	CTCCAGGGGATCAAGTATTGTCAACTTAAGAGACTTGTGTTCAACACCTGGATC	2173
Db	2066	GAGCTCTAGTGTGGAATCTCTGTTTAAACCATCGAACTTTCGTTTAAAGCGTGGAG	2125
Qy	2174	AGAGCTGAGCTATAGATGATGTTTGTATTGAAGAAGAAACACAGAGGAGGAGGTGG	2233
Db	2126	AGAAAGAGGTACACAGTCAAGTTTGTAGCAGAAAGGAGTATGACGACACAGG	2185
Qy	2234	CTAGCTTTG 2242	
Db	2186	CTGAGTTTG 2194	
LOCUS	AY142613	2296 bp	linear
RESULT 8			PLN 23-SEP-2002
LOCUS	AY142613		

SAVRNASFLIYAAGTAGMATRATVATYKVCWSTCGFSGDILAAVDRAILDGVVLSL
SLGGSPAYVDRDIALGAVSLYSGVGMGTPLLELVNKGNSSSNLCPLQSLDSSIVRG
RDPAPANLGNRGLTAVSLYSGVGMGTPLLELVNKGNSSSNLCPLQSLDSSIVRG
KIIVCDRGNARVEKAVVDAGGLIMANTAASEELVADSHLLPALAVGKKTGDL
LREYVSDSKPTALLVFGTLDVDFKPSVVAHPSRGPENTVPEILKEDVPDGPVAIL
AGWSDAIGPTDKDSDRRTOFNMISGSCPHI SLAGLLKAAHPSPSPSAIKALIM
TTAYVLNDNTNAPLHDAADNSLNPHYAGSHVDPOKALSPGLVYDISTEYIRFLCSL
DYTVDSHIVAIKRPSPVNCCKFSDPQLANVPSFVLFPGKRVVRYTREVTVNGAASSV
KYTVNAPSAGVLSVAPSKLSFKSVGEKKRYTVTVFSKKGVSMTNKABFSGITWSPQ
HEVRSVPASWNRP"

3'UTR

2866..2296

/gene="Ac2g05920"

BASE COUNT 532 a 515 c 546 g 703 t

ORIGIN

Query Match

Best local Similarity 55.3%; Pred. No. 8.5e-126; Length 2296;

Matches 1033; Conservative 0; Mismatches 162; Indels 24; Gaps 4;

QY	380	ACTCGGTGATGCTCTAAATCTCGGTTTGGTCAAGGCACAAATATTCGGCGTCTTGATA	439
Db	344	AATTCGGTGTTCAGATCTCGGTTCTTCTTAACGGCGTTATCATCGGAGTTTATAGATA	403
QY	440	CTGGAGTTTGGCCTGAAAGTCTAGCTTTGACGATACCGGAATGCCCTTCGATTCACGGA	499
Db	404	CTGGCGTTTGGCCTGATCTAGAGCTTCGATGATCTGATATGCTGAGATTCCTTCTA	463
QY	500	AATGGAAGGATTTGCGCAAGAGGAGAAAGTTTCAGTTCTTCGAGCTGTAACCGGAAGC	559
Db	464	AATGGAAGGAGATGTGAATCTGTTTCCGATTTCCGATTTCCGATTTCCGATTTCCGAT	523
QY	560	TAATCGGTGTAGATTTCTCATCAGAGGACACGCTGTCGCTTAATTCACGAGAGGAATCAC	619
Db	524	TTATCGGAGCTAGAGCTTCTCCAAAGATTTCAAATGGCTTCTGTTGGTGGTGGTTTCGA	583
QY	620	CAAAATGCTCTGCTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCG	679
Db	584	GTAA-----CGGTGAATCTGTTTCTCTCTGCTGATGTTGACGGACATGGAACATATT	637
QY	680	CCTCAACAGTTGGTGGATCTCTGTTTCGATGCGGAATGTTCTTGGCAATGGAGCTGGTG	739
Db	638	CAACTACCGCGCGGATCCGCGCTTAGAAACGCTAGCTTCTCGGTTACGCGCGCGGTA	697
QY	740	TGGCTCGTGGATGGCTCTGGAGCTCACATTGGAGTCTATAAGTCTGTTGTTCAATG	799
Db	698	CGGCGAGAGGTATGCCACTCGTCTCGTGTGCTACTTATAAAGTTTGTGGAGTACTG	757
QY	800	GTTGTTACAGCTCTGACATTTAGCAGCTATAGATGTAGCGATTCAAGATAAAGTCCGATG	859
Db	758	GTTGTTTGGATCTGATATCTAGCTGCTATGATCGAGCTACTTATGATGGTGGTGGATG	817
QY	860	TTCTTTCCGCTTCCGCTTGGCGTTTCCCTATTCCTTTTGTATGATGACACAATCGCCATG	919
Db	818	TGCTTTGCTTATCTCTTGGTGGTGGTCTGCTGCTGCTATTTACGGATACGATTGCGATTG	877
QY	920	GAAATTCGAGACCATGGAACCGGTATATCTGTAATCTGCTGAGCTGGTGAACACCGTTC	979
Db	878	GAGCGTTTTCGGCTATGAGAGAGAGTGTGTTTGTGTCTGTGTTCTGCTGGTAAATAGTGGT	937
QY	980	CAATCGAAGCTCTGTTCCAAACACAGCTCTCTGGTCTCAACATTTGGCGCAGGCACGC	1039
Db	938	CTACTAGAGCTTCTGTTCCCAATGTTGCTCTCTGGTGTATGACTGTTGTTGCTGTTACTT	997
QY	1040	TTGATCGAAGATTTCCGCTGTGTCAGATTAGCCAAACGGAAAGCTTCTCTATGGAGAGT	1099
Db	998	TAGATAGAGATTTTCGGCTTTTGGGATCTCGGTAAACGGGAAACGACTTACCGTGTGTT	1057
QY	1100	CATTGATTCGGGAAAGGTATAAAGATGCGGGAGAGAGTTGAGGTGATTTACGTCA	1159
Db	1058	CGCTGATAGCGGTGTAGGAATGGGA-----CGAAGCCGCTTGAATTTGTTTATAATA	1111
QY	1160	CAGGAGGAGATAAAGGAGTGTGTTCTGTTGAGAGGCTCACTTCCAGAGAGGAATCC	1219

Db	1112	AAGGGAATAGTAGTTCAGTAATCTTTGTTTACCTGTTGCTTGTGATTGAGTATTGTTT	1171
QY	1220	GAGCAAAATAGTGAATTTGTGATCGCGAGTCAATGAAGATCGGAGAAAGAGACCGG	1279
Db	1172	GTGGGAAGATTGTTGTTTGTGATAGAGTCTTAATGCTAGAGTTGAGAAAGAGCTGTGG	1231
QY	1280	TTAAAGAGCTGGAGGAGTTGCAATGATCTTAGCCATACAGAGATCAACCAAGAGAAG	1339
Db	1232	TTAGAGATGCTGCTGGTGTAGGAGATGATAATGCGAATCTGCTGCGAGTGGAGAGAGC	1291
QY	1340	ATTCTATTGAGCTTCACTCTTACCAGCTACATTTGATTGTTTACATGAGTCACTGCTTC	1399
Db	1292	TTGTGGCGGATGATCATTTTCTTCCCGGATCGCTGTAGGGAAGAGACTGGTGTATTAC	1351
QY	1400	TGAAGGCTTATGTTTAATGCCACCGTGAAACAAAGGCGCGGATAATTTTGTGGTGGTACGG	1459
Db	1352	TTAGGAGTATGTTTAAGTCAGATTTCTAAACCAACCGCTCTTCTGTTTAAAGGAACCG	1411
QY	1460	TGATGGAGGTACAGAGCACCAGGAGTGGCTAGTTTTCAGCTCGAGGACCGAGTTTAC	1519
Db	1412	TTCTTGAGCTTAAAGCCGCTCTCTGCTGTTTGTGCTTTTGTGCTGAGAGGCTCCTAATCTG	1471
QY	1520	CCATCTCTCGATACTAAACCGGATATGTTGCTCCGGAGTCAATATCATTCATTCGGCTT	1579
Db	1472	TTACTCTCTGAAATCTTGAAGCTCTGATTTATTTGCTCTGAGTTAAATTTTGTGCTGGTT	1531
QY	1580	GGCTTCAAAATCTAGGACCAACCGGACTTCTTATGATTCAAGAAGAGTTTAACTTCACTG	1639
Db	1532	GGTCTGAGCTTATTTGTTCTTCTTCTGACAGGACTCTAGGAGGACTCAGTTCAACA	1591
QY	1640	TAATGTCAAGAACTTCAATGTTCTTGTCCACATGTTAGCGGATCAGTCTCTTATTCGGT	1699
Db	1592	TCATGTCAAGTACGTCAATGTCTGCTCCACACATCAGTGGTTTACCGGCTCTTTTGAAG	1651
QY	1700	CTGCATACCCGAACTGGTCTCCAGTCAATCAATCGCATTCGATGATGACACAGCGATT	1759
Db	1652	CAGCTCACCTTGAGTGGAGTCCGAGTCTATCAATCAGCTCTCATGACTACAGCTTACG	1711
QY	1760	TGTCAGATCGTCAAGGGAAGCGATTAAGGATGT-----AACAAACCGACCGGCTGTGT	1813
Db	1712	TTCTTTGACAAACACCAACCGCTCTCTCCATGATGCTGACAGACAAACAGCTATCTAACCCAT	1771
QY	1814	TTGCTGATTGGAGCAGGCGATGTGAATCCGAAAAAGGCGATAAACCGGAGTTGGTTTACA	1873
Db	1772	ATGCTCAAGCTCGGCTCGGCGCATGTAGATCCCAAAAGGCTCTCTCACCAGCTCTTGTCTACG	1831
QY	1874	ACATCAACAGTGTGATTAATTAATCTTACCTCTGCACTCTTGGATTCAAGATCAGATA	1933
Db	1832	ACATCTCAACCGAGGAATACATCAGGTTTTTGTGCTCTCTAGACTACACAGTCCGATACA	1891
QY	1934	TTTTAGCAATCACTCATTAAGACGTGAGCTGCAATGGAATTTGCGGAAAAACCGCGGTT	1993
Db	1892	TTGTTTGGGAT-----TGTGAAGCGACCTAGCGTTAATGCTCGAAGAGTTCTCAGATC	1945
QY	1994	TTAGTCTCAATTAACCGCTCGATAGCGGTGATTTTCAACCGTGGCAAGACTACGGAGATGA	2053
Db	1946	CTGGTCACTCACTACCCCAAGTTTCTCGGTTTTTGGGGTAAAAAGAGTTGTGCGGT	2005
QY	2054	TCACAAGCGGTGTCACTAAGCTTGGAGTCTTAATCGATATCTCAGTGAATGTCAAGG	2113
Db	2006	ACACTCGGGAAGTAAACAAATGTTGGTGCAGCAAGCTCGGTTTTTACAAAGTACGGTTAATG	2065
QY	2114	CTCCAGAGGGGATCAAGATTATTTGCAATCTTAAGAGACTTGTGTTCAACACAGCTGGATC	2173
Db	2066	GAGCTCTAGTCTCGGAATCTCTGTTAAACCATCGAACTTTCGTTTTTAAAGCGTGGAG	2125
QY	2174	AGACCTGAGCTATAGAGTATGTTTGTATTGAAGAGAAAAACAGAGAGGAGGAGGTGG	2233
Db	2126	AGAAGAGAGGTACACAGTCCAGCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT	2185
QY	2234	CTAGCTTTG 2242	
Db	2186	CTGAGTTTG 2194	

```

RESULT 9
LOCUS AY035090
DEFINITION Arabidopsis thaliana putative subtilisin serine protease
ACCESSION AY035090
VERSION AY035090.1
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi;
1 (bases 1 to 2541)
Yamada, K., Liu, S. X., Sakano, H., Pham, P. K., Banh, J., Chung, M. K.,
Goldsmith, A. D., Lee, J. M., Quach, H. L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R. W., Ecker, J. R. and Theologis, A.
Arabidopsis Full Length cDNA Clones
Unpublished
2 (bases 1 to 2541)
Yamada, K., Liu, S. X., Sakano, H., Pham, P. K., Banh, J., Chung, M. K.,
Goldsmith, A. D., Lee, J. M., Quach, H. L., Tang, C. C., Toriumi, M., Yu, G.,
Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,
Kim, C., Koesema, S., Lam, B., Lin, J., Meyers, M. C., Miranda, M.,
Narusaka, M., Nguyen, M., Palm, C. J., Sakurai, T., Satou, M., Seki, M.,
Shinn, P., Southwick, A., Shinozaki, K., Davis, R. W., Ecker, J. R. and
Theologis, A.
Direct Submission
Submitted (03-MAY-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN
Arabidopsis Full-Length cDNA'); Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.
The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAPL cDNAs: Yamada, K., Liu, S. X.,
Sakano, H., Pham, P. K., Banh, J., Chung, M. K., Goldsmith, A. D.,
Lee, J. M., Quach, H. L., Tang, C. C., Toriumi, M., Yu, G., Bowser, L.,
Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E.,
Lam, B., Lin, J., Meyers, M. C., Miranda, M., Nguyen, M., Palm, C. J.,
Shinn, P., Southwick, A., Davis, R. W., Ecker, J. R. and Theologis, A.
Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)
contributed equally to this work as PIs.
Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to GenBank.
FEATURES
Location/Qualifiers
1..2541
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAF109-15-A05 (R09230)"
/note="This clone is in a modified pBluescript vector
(FLC-1) as a BamHI/XhoI insert.
ecotype: Columbia"
1..2541
/gene="At2g05920"
1..22
/gene="At2g05920"
misc_difference 2
/gene="At2g05920"
/note="compared to genomic sequence"
/replace="c"
23..2287
/gene="At2g05920"
/codon_start=1
/evidence=experimental
/product="putative subtilisin serine protease"
/protein_id="AAK595.1"
/db_xref="GI:14334834"
translation="MASSSSSSSSIIITTFLLHHTAKKTYIRVNHDKPSFL
THDWTQLNSESLLYTTSPHFGAYLDSTADSLSSNSILDI FEDPLYTLH
TTRTFELGLNSEFGVHDLGSSNGVIGLDTGWPESRFPDDTDPMEIPEISKWGECE
ESGSDFKLKNKLIAGRSFKGFOMASGGGSSKRESVSPRDVDGHTGHTSTTAAG
SAVRNASFLYAGTARGATRAVATYKVCWSTCCFSDILANDRAILLDQVDVLSL
SUGGSAPYRTDITIAFGASMERGVFVSCSAGNSGPTASVANVPMWMTVGAGTLG
RDFPAPANGKRLTGVSLYSGVMGTPLKLELVYKNKNSSSNLCIPGSLDSTVIRG
KIVDCRGVNAVEKGVADAGLGIMTAMTAASBEELVADSHLLPAI AVGKTKGDL
LREYKSDSKPTALLAFKGTLDVKPSVFAAFSSRGNPTVPELTKDVTGPGVNIIL
AGWSAIDGPTGLDKSRTOFNMSGTSVCPHISGLAGLLKAAHPESPSAISKALM
TAYVDNTNAPLHDAANSJNPAHSGHVDPOKALSPGLVDYDISEEYIRFUCSL
DYTVDHIVAIKRPVSNCKSPFQNLNPSFVFGKRVVRYTRVTVNVGAASSV
YKVTVNGAPSVGISVSKPSKLSFKSVGEKKRYTVTFVSKKGVSMTNKAEFGSITWNPQ
HEVRSPVAFSWNRF"
2288..2541
/gene="At2g05920"
misc_difference 2524..2525
/gene="At2g05920"
/note="not present in genomic sequence"
BASE COUNT 617 a 549 c 600 g 775 t
ORIGIN
Query Match 20.7%; Score 481.8; DB 8; Length 2541;
Best Local Similarity 55.3%; Pred. No. 8.6e-126;
Matches 1033; Conservative 0; Mismatches 812; Indels 24; Gaps 4;
QY 380 ACTCCGGTGTATGGTCTAAATCTCGTTTGGTCAAGGCACACATTATCGCGTGTCTGATA 439
Db 366 AATTCGGTGTTCACGATCTCGGTTCTTCTTAAAGCGGCTTATCATCGGAGTTTAGATA 425
QY 440 CTGGAGTTTGGCTGAAGTCTAGCTTTGACGATACCGGAATGCTTCGATTCACACGA 499
Db 426 CTGGCGTTTGGCGTGAATCTAGAGTTCGATGATGATGATGATGATGATGATGATGAT 485
QY 500 AATGAAAGGGATTTGCCAAGAGGAGAAAGTTTCAGTCTTTCGAGCTGTAAACCGGAAGC 559
Db 486 AATGAAAGAGAGATGTGAATCTGGTTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 545
QY 560 TAATCGGTCTAGATTTCTATCAGAGGACACCGTGTCTGCTTAATTCACAGAGGATCAC 619
Db 546 TTATCGGAGTAGAAGCTTCTCCAAAGGATTTCAAAGGATTTCAAATGCTTCTGTTGTTTTCGA 605
QY 620 CAACAATGCTGTGATACATTTCCGCAAGAGATTCACGGGACACGGGACTCACACCG 679
Db 606 GTAA-----GGGTGAATCTGTTTCTCTCGTGTGTTGACGAGATGGAACACATACAT 659
QY 680 CCTCAACAGTTGGTGTGATCTCTGTTTCGATCGGAATGTTCTTGGCAATGGAGTGTGTG 739
Db 660 CAACACTACCGCGGAGATCCGCGTTAGAAACGCTAGCTTCTCGTTACGCGCGGCTA 719
QY 740 TGGCTCGTGGGATGGCTCCTCGGAGCTCACATTCGAGTCTATAAAGTCTGTTCGTTCAATG 799
Db 720 CGGCCAGAGTATGGCCACTCGTGTCTCGTGTGTTGCTACTTATAAGTTTCTGGAGTACTG 779
QY 800 GTTGTGTACAGTCTGATCATTCATGACAGCTATAGATGTAGCGATTCAAGATAAGTCGATG 859
Db 780 GTTGTGTGATCTGATATCTAGCTGTATGCTATGATGCTATGCTATGCTATGCTATGCTATG 839
QY 860 TTCTTTTCGCTTTCCCTTGGCGGTTTCCCTATTCTTGTATGATGACACAAATCGCATG 919
Db 840 TGTCTTTCGTTATCTCTTGTGTGGTGTTCGTCGGTATTATCGGATACGATTGCGATTG 899
QY 920 GAACATTCGAGCCATGGAAACGCGGTATATCTGTAATCTGTGACGCTGTTAAACAAACGGTC 979

```

900 GAGCGTTTTCGGCTATGAGAGAGGTTGTTTTTGTGTTCTGTTCTGCTGTAATAGTGGTC 959
980 CAATCGAAAGCTCTGTTGCAACACACAGCTCTCTGGGTTCTCAACATTTGGCGCAGGACGC 1039
960 CTACTAGAGCTTCTGTTGCCAATGTTCTCTCTGGGTTATGACTGTTGGTCTGGTACTT 1019
1040 TTGATCGAAGATTTCCCGCTGTGTCAGATTAGCCACCGAAGAGCTTCTCTATGGAGAGT 1099
1020 TAGATAGAGATTTTCGGCTTTTTCGGAATCTCGGTAACCGGAACGACATACCGGTGTT 1079
1100 CATTGTATCCGGGAAAGGTATAGAAATGCCGGGAGAGAGGTGAGGTGATTACGTCA 1159
1080 CGCTGTATAGCGGTGTAGGAATGGGA-----CGAAGCGGCTTGAATTTGTTTATAATA 1133
1160 CAGGAGAGATTAAGGAGTCACTTCTGTTTGAAGAGGTCACCTTCCAGAGAGAATCC 1219
1134 AAGGGAATAGTAGTTCGAGTAATCTTTGTTTAACTCGCTTGTGATTCGAGTATTGTTTC 1193
1220 GAGCAAAATGGTATTGTTGATCGCGAGTCAATGGAAGATCGGAAAGGAGAGAGCGG 1279
1194 GTGGGAAGATTGTTTGTGATAGAGGTGTTAATGCTAGATTGAGAAAGGAGCTGTGG 1253
1280 TTAAGAGAGCTGGAGGAGTTGCAATGATCTTAGCCATACAGAGATCAACCAAGAGAAG 1339
1254 TTAGAGATGCTGGTGGTTAGGATGATAATGGCGAATCTGCTCGAGTGGAGAGGAGC 1313
1340 ATTCTATTGAGTTCATCTCTTACAGCTACATGATTGTTTACCTGAGTCACTGCTTC 1399
1314 TTGTGGCGGATGATCAATTTGCTTCCCGGATCGCTGTAGGGAAGAGACTGGTGAATTAC 1373
1400 TGAAGGCTTATGTTAATGCCAGCTGGAACCAAGGCGCGATAATTTTGGTGGTACGG 1459
1374 TTAGGAGTATGTTAAGTCAGATCTTAACCAACCGCTCTTCTGTTTAAAGGAGAGG 1433
1460 TCATTGGGAGTTCAGGACCGGAGTGGTTCAGTTTTCAGCTCGAGGACCGAGTTTAG 1519
1434 TTCITGAGTTAAGCGCTCTCTGTGTTGCTGCTTTTAGCTCGAGAGTCTTAATACTG 1493
1520 CCAATCTTCGATCTAAACCGGATGATGTTCTCGGAGTCAATATCATTTGCGGCTT 1579
1494 TTACTCTCGAATCTGAAGCTGATGTTATTGTTCTCGGAGTTAATATTGCTGGTGT 1553
1580 GGCCTCAAAATCTAGGACCAACCGGACTTCTTATGATTCAGAGAGTAACTTCACTG 1639
1554 GGTCTGAGCTATTGGTCTCTACTGCTTGAAGAGGACTCTAGGAGACTCAGTTCACCA 1613
1640 TAATGTCAGGAACCTCAATGCTTGTCCACATGTTAGCGGAATCACTGCTTCTTATCGGT 1699
1614 TCATGTCAGGTACGTCAATGTCATGCCACACATCAGTGGTTAGCGGCTTTTGAAG 1673
1700 CTGCATACCCGAACTGGTCTCCAGTGGCAATCGCAATCGCATTTGATGACACACGCGATT 1759
1674 CAGCTCACCTGAGTGGAGTCCGAGTGTATCAATCAGCTCTCATGACTACAGCTTAAG 1733
1760 TGTAGATGCTCAAGGGAAGAGTAAGGATGTT-----AACAAACCGCGGTGTGT 1813
1734 TTCTTGACACACCAACGCTCTCTCCATGATGCTGCAGACACAGAGCTATCTAACCCAT 1793
1814 TTGCGATTGGAGCGGCGATGATGATCGCAAAAGGCGATAAACCAGGATTTGTTTACA 1873
1794 ATGCTCAGCGTTCGGGCCCATGATGATCCCAAAAGGCTCTCTACAGGCTTTGTTCTAG 1853
1874 ACATTCACCAAGTGGATTACATACTTACCTCTGCTGCTGCTGTTGATTTCACAGATCAGATA 1933
1854 ACATCTCAACCGAGGAATACATCAGGTTTGTGCTCTCTAGACTACACAGTCCATCACA 1913
1934 TTTTAGCAATCACTCATAGAACGTCGTCGAATGGAATATTCGGGAAACACCGGGT 1993
1914 TTGTTGCGAT-----TGTGAAGCGACCTAGCGTTAACTGCTCGAAGAGTTCTCAGATC 1967
1994 TTAGTCTCAATACCGTCTGATAGCGGTGATTTTCAACGTTGGCAAGTACGAGATGA 2053
1968 CTGGTCAGCTCACTACCCAGTTTCTCGGTTTTGTTGGGGTAAAGAGTTGTCGGT 2027

QY 2054 TCACAAGCGGTGTCACATAACGTTGGGAGTCTTACTCGATATACTCGATGAATGTCAAGG 2113
Db 2028 ACATCGGAGTAGTACAAATGTTGTCGAGCAAGCTCGTTTACAAAGTGACGGTTAATG 2087
QY 2114 CTCAGAGGGGATCAAAAGTTAATGTTCAATCTTAAGAGACTTGTGTTCAACACGTGATC 2173
Db 2088 GAGCTCTAGTGTGCGAATCTCTGTTAAACCATCGAAACTTTCGTTTTAAAGCGTGGAG 2147
QY 2174 AGACGCTGAGCTATAGAGTATGTTTGTATTGAGAGAGAAAACAGAGGAGGGAGGTGG 2233
Db 2148 AGAAGAGAGGTACACAGTCACTGTTTGTAGCAAGAAGGAGTGATGATGCGAACAGG 2207
QY 2234 CTAGCTTTG 2242
Db 2208 CTGAGTTTG 2216

RESULT 10

BT001082

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

BT001082 2274 bp mRNA linear PLN 18-OCT-2002
Arabidopsis thaliana At5g67360/X8K14_8 mRNA, complete cds.

BT001082

BT001082.1

FLI CDNA.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 2274)

Chen, R., Chen, H., Kim, C.J., Shinn, P., Bowser, L., Carninci, P.,

Chan, M.M., Chang, C.H., Dale, J.M., Havaehizaki, Y., Hsuan, V.W.,

Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,

Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,

Pal, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,

Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yamada, K.,

Yu, G., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. and

Ecker, J.R.

Arabidopsis ORF clones

Unpublished

2 (bases 1 to 2274)

Chen, R., Chen, H., Kim, C.J., Shinn, P., Bowser, L., Carninci, P.,

Chan, M.M., Chang, C.H., Dale, J.M., Havaehizaki, Y., Hsuan, V.W.,

Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,

Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,

Pal, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,

Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yamada, K.,

Yu, G., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. and

Ecker, J.R.

Direct Submission

Submitted (18-OCT-2002) Salk Institute Genomic Analysis Laboratory

(SIGNAL), Plant Biology Laboratory, The Salk Institute for

Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

USA

The Salk, Stanford, PGEC (SSP) Consortium members constructed and sequenced the PUN1 (ORF) clones using the RAFL cDNAs: Cheuk, R., Chen, H., Kim, C.J., Shinn, P., Ban, J., Bowser, L., Chan, M.M., Chang, E., Dale, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Davis, R.W., Theologis, A., and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES		Location/Qualifiers	
source		1..2274	
		/organism="Arabidopsis thaliana"	
		/mol_type="mRNA"	
		/db_xref="taxon:3702"	
		/chromosome="5"	
		/clone="U21848"	
		/note="This clone is in puni 51"	
		ecotype: Columbia	
CDS		1..2274	
		/note="cucumisin-like serine protease (gb AAC18851.1)"	
		/codon_start=1	
		/product="At5g67360/k8k14_8"	
		/protein_id="AA04863.1"	
		/db_xref="GI:24111425"	
		/translation="MSSSFLSSSTAFPLLCLGRCHVSSSSDQGTIVHMAKSMOPSS FDLSHNYDSLSRLSISDAELLYTENAIFGSTRLOEADSLMQPQVIVLPBHR YELHTTLPFLDEHDTADLPFAGSYSDVYVGLDTGVMPESKYSDBGFPIPS WKGCEAGTFTASLCNKRKLIAGFFARGYESTMGPIDESKSRSPRDDHGHHTSS TAAGSVVGGASLLYAGSGTAGMAPRAVAVKVMGLGCFSDILLAAIDKADLVN VLSNLSGGMSDYVDGVAICAFAPAMERGIIVSCAGNAGPSSSSISNVAPWITTVGA GLDRDPALAILNGKNFTGVSLFKGEALPDKLLPIFYAGNASNATNGNLMTGTLI PEKVKIKVMDRGINARVQKDVVKAAGVGMILLANTAANGBELVADAHLLPATTG EKAGDIIRHYITDPNPNTASISILGTGVGKPPVAAVFSRGPNSITPNILKPDIIA PGVNIILAACTGAOPTGLASRRVEFNIISGTSMSCPHVSGLAALLKSVHPRCSFAA IRSLMTAYTKYKDGKPLLDIAFGKSTPFDHGHVSPPTATNPGLIYDLTETVSL GFLCALNYTPQIRSVRRNYTCDFSKSYSDADLNYPSPFVNVGNGVAYKYTRTVSV GGAGTYSVKVTSSETTGKISVEPAVLNFKENAKKSYITVTFVDDSKPSSNSFGSIE MSDGKVVVGSFVA:SWT"	
BASE COUNT		535 a	566 c 571 g 602 t
ORIGIN			
Query Match		20.6%; Score 480.4; DB 8; Length 2274;	
Best Local Similarity		56.2%; Pred. NO. 2.1e-125;	
Matches 973; Conservative		0; Mismatches 741; Indels 18; Gaps 3;	
QY	210	ACTTCTCTACTCTCTGCTCTGCGATTGAAGGATTTGCTGCTCAGTTGACTGAATCAGA	269
DB	192	ATTGCTCTACACTTACGAGACGGGATTCATGGATTCTCAACACGGCTAATCAAGAAGA	251
QY	270	AGCCGAGATCTGAGATATTACCTGAAAGTTGTTGCACTGAGACCTGACCACTGTTCTCA	329
DB	252	AGCCGACTCGCTCATGACTCAACCTGGTGTATCTCGTTTTACGGGACACCGTTACGA	311
QY	330	GGTTCAAAACCACTTACTCTTACAGTTCTTGGGACTCGACCGGTTTTGGAACTCCGGTGT	389
DB	312	GCTACACACCACTCGTACTCTCTCTCTCGGCTCTCGACG---AACACACCGCAGATCT	368
QY	390	ATGGTCTAAATCTCGTTTGGTCAAGGCACATTAATCGGCGTCTTGATATCTGGAGTTG	449
DB	369	GTTCCTGAAGCCGCTCTTACAGCGACGTGCTGCTGGAGTTCTCGATACCGGAGTTTG	428
QY	450	GCCTGAAAGTCTAGCTTTGACGATACCGGATGCTTTCGATTCCACGGAATGAAAGG	509
DB	429	GCCGAGAGTAAAGCTACTCCGACGAGGATTCGGTCTTATCTCTCTCTCTGAAAGG	488
QY	510	GATTTGCCAAGAGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAAGCTAATCGGTGC	569
DB	489	CGGATCGGAGGCGGAAACCAATTTACCGCTTCTCTGTAAACCGTAAACTAATCGGAGC	548
QY	570	TAGATTCTTCATCAGAGACACCGTGTGCTTAATTCACAGAGGATCAACCAACATGCC	629
DB	549	AAGATTCTTCGCTCTGTTTAC-----GAATCAACGATGGGACCAATCGATGAATC	599
QY	630	TCGTGAATACATTTCCGGAAGAGATTCAACGGGACACGGGACTCACACCGCTCAACAGT	689
DB	600	TAAAGAACTCAAGATCTCTTAGAGACGACGACGGTCACCGAATCAACCTCATGACCGC	659
QY	690	TGGTGATCTCTGTTTCGATCGCGAATGTTCTTGGCAATGAGCTGGTGTGGTCTGTTGG	749
DB	660	CGCTGGATCCGTCGTTGAAGGAGCTAGCTTATTAGGTACGCTTCAGGAAACAGCTCGTG	719
QY	750	GATGCTCTCGGAGCTCACATTCGAGCTATAAAGTCTGTTGGTTCAATGGTTGTTACAG	809
DB	720	TATGCTCCACCGCGCTCGTGTGCTGTTTCAAAAGTCTGTTGGCTCGGTGTTGTTTCAG	779
QY	810	CTCTCACATTTCTAGCAGCTATAGATTAGCATTCAAGATAAAGTCGATGTTCTTCGCT	869
DB	780	CTCAGATATTTTAGTGTCAATCGATAAAGCCATCGCGATTAACGTCAATGCTCTGTGAT	839
QY	870	TTCCCTTGGCGGTTTCCCTATTCCTTTGTATGATGACACAATCGCCATCGGAACATTCGG	929
DB	840	GTCACTCGGCGTGGAAATGTGGATTAATATAGACACGGTGTTCGATCGGAGCATTCGC	899
QY	930	AGCCATGAAACCGCGGTATATCTGTAATCTGTGCAGCTGGTAACAACGCTCCAATCGAAAG	989
DB	900	CGCCATGAAAGAGGAGATTTTAGTATCTTGTCTAGCTGGTAATGCTGTGCGAGTAGTTC	959
QY	990	CTCTGTTGCAACACACAGCTCCTTGGGCTCTCAACCAATTCGCGCAGCAGCGTTGATCGAAG	1049
DB	960	TAGTTTATCAACAGTAGTCTCAATGATCACAACCTCTTGGTCTGTACTTTAGATCGTGA	1019
QY	1050	ATTTCCTCGCTGTGTGTCAGATTAGCCACCGAAAGCTTCTCTATGAGAGTCAATGTATCC	1109
DB	1020	TTTTCCGGCGCTTGGGATTCCTCGGTACCGCAAGAAATTTCAACCGAGTTCTTGTGTTAA	1079
QY	1110	GGGAAAAGGTATAAAGAATGCGCGGAGAGAGGTTGAGGTGATTTACGTACAGAGGAGAGA	1169
DB	1080	AGGAGAAGCTTTTACCGGATAAATTTCTGCCGTTTATTTACGCTGGGAATGCTAGTAATGC	1139
QY	1170	TAAAGGAGTGTGTTCTTTTGTAGAGGCTCACTTCCAGAGAAGAAATCCCGAGGCAAAAT	1229
DB	1140	TACTAATGTTAATCTCTGTATGACCGGAACCTTTGATCCCGGAGAAAGTAAGGGGAAGAT	1199
QY	1230	GGTGTATTTGTGATCGCGGAGTCAATGGAAGATCGGAGAAAGAGAGAGCGGTTTAAAGAAGC	1289
DB	1200	TGTGATGTGTATAGAGGAATTAATGCTAGAGTTTCAAGAAAGGTGATGTGTTTAAAGCAGC	1259
QY	1290	TGAGAGGTTTGAATGATCTTAGCCAAATACAGAGATCAACCAAGAGAAGATTTCTATTGA	1349
DB	1260	AGGTGAGTTTGAATGATTTCTGGCTAAACACTCGCGCAATGTTGAGAGCTTTGTTGCGGA	1319
QY	1350	CGTTCTATCTTTACCACTACATTTGTTGTTTACACTGAGTCAGTCCCTTCTGAAGCTTTA	1409
DB	1320	TGCTCATTTTGTACCGCGACACCGGTTGGTGAAGAAAGCCGGTGATATAATCCGGGATTA	1379
QY	1410	TGTTAATGCCACGGTGAAACCAAGCGCGGAGTAATTTTGGTGTGATCGGTGATTTGGGAG	1469
DB	1380	CGTGACTACTGATCTTAATCCACCGCTTCGATTTCAATCTTTAGGAAACAGTCTGGTGT	1439
QY	1470	GTCAACGAGCACCGGAGTGGCTCAGTTTTCAGCTCGAGACCGAGTTTAGCCAAATCCCTTC	1529
DB	1440	TAAACCATCTCGGTTGTGCGACGCTTAGCTCACGTGACCGGAATTCGATTACACCGAA	1499
QY	1530	GATCTAAACCGGATATGATTTGCTCCGGGAGTCAATATCATTTGCGGCTTGGCTCAAAA	1589
DB	1500	TATTTTAAACCGGATCTGATCGCTCTGGAGTAAACATCTCTCGCGGCTGGACCGGTGC	1559
QY	1590	TCTAGAACCAACCGGACTTCTTATGATTCAGAGAGTTAACTTCACTGTAATGTCAGG	1649
DB	1560	TGCTGACCAACCGGACTCGCTTCGATTTCTCGCGCGTGGAGTTCAATATCATCTCTGG	1619
QY	1650	AACTTCAATGCTTTGTCCACATGTTAGCGGAATCACTGTCTTATCCGGTCTGTCATACC	1709
DB	1620	GACCTCGATGTTCTTGCCCTCAGTTAGTGTGTTTAGCGGCTTCTCAAGTCGGTGCATCC	1679
QY	1710	GAACTGCTTCCAGCTGCAATCAAAATCCGATTTGATGACAAACAGCGGATTTGTAGATCG	1769
DB	1680	TGAATGTAGCCCGCGCGATTAGATCGGCGCTTATGACCAACCGCTTACAAAACCTTACAA	1739
QY	1770	TCAGGGAAGCGATA-----AAGATGTTAACAACACCGCGGTGTCTTTGCGATTGCG	1823
DB	1740	AGACGGTAAACCGTTTACTCGACATCCGACAGGAAGCCCTTCGACGCCGTTGATCACCG	1799
QY	1824	AGCAGGCAATGTGAATCCCGCAAAAGCGGATAAACCGGAGTTGTTTACAAATTCATCAAC	1883

DB 1800 TGCAGGACAGTGTCAACAAACAATGCCACTTAATCCAGGACTCATCTACGATCTTAAGAC 1855

QY 1884 AGTGATTACATAACTTACCTCTGCACCTCTTGGATTCAACAAGATCAGATATT 1935

DB 1860 GGAGATTACTTAGCTTCTCTGCGGATTGAATTACACATCGCGCAATT 1911

RESULT 11

AY0911773 2520 bp mRNA linear PLN 13-APR-2002

LOCUS Arabidopsis thaliana AT5G67360/K8K14_8 mRNA, complete cds.

AY0911773

ACCESSION AY0911773.1 GI:20147210

VERSION

KEYWORDS FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 2520)

AUTHORS Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE Arabidopsis cDNA clones

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2520)

AUTHORS Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE Direct Submission

JOURNAL Submitted (26-MAR-2002) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Banh, J., Bowser, L., Chang, E., Dale, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A., and Ecker, J.R.

Shinn, P. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Banh, J., Bowser, L., Chang, E., Dale, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A., and Ecker, J.R.

QY 630 TCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCCCTCAACAGT 689
Db 631 TAAAGAACTCAGATCTCTAGAGACGACGAGCGGTCAAGGAACTCACACCTCATCGACCGC 690
QY 690 TGGTGGATCCTCTCTTTCCGATGGGAATGTTCTTGGCAATGGAGCTGTGTGCTGCTGG 749
Db 691 CGCTGGATCCGCTGTTGAAGAGAGTACGTTATTAGGCTAGCGTTCCAGGAACAGCTGCTGG 750
QY 750 GATGCTCCTGGAGCTCACATTCGAGTCTATAAAGTCTGTGTGGTTCAATGGTTGTACAG 809
Db 751 TATGGCTCCAGCGCTCGTGTGCTGTTTACAAAGTCTGTGTGCTGGTGTGTTTCAG 810
QY 810 CTCTGACATTCAGAGCTATAGATGAGGATTCAGGATTCAGAGTAAAGTGCATGTTCTTCGCT 869
Db 811 CTCAGATATTTAGCTGCAATCGATAAAGCCATCGCCGATAAGCTCAATGCTCTGCTCAT 870
QY 870 TTCCTTTGGCGTTTCCCTATTCCTTTGTATGACACAAATCGCCATTCGACATTCG 929
Db 871 GTCACTCGCGGTGGAATGTCGGAATATTATAGAGCGGTGTTCGATCGGAGCATTCGC 930
QY 930 AGCCATGGAACGCGGTATCTGTAATCTGTGACGCTGTTAAACAAACGCTCCAATCGAAAG 989
Db 931 CGCATGGAAGAGGATTTTAGTATCTGTCTGACGTGTTAATGCTGTCCGAGTAGTTC 990
QY 990 CTCTGTTGCAACACAGCTCCTCGGTCTCAACCATTCGCGCAGGACGCTTGAATCGAAG 1049
Db 991 TAGTTTATCAACGCTAGCTCATGATCAACAACTGTTGTGCTGTGTTAGATCGTGA 1050
QY 1050 ATTTCCCGCTGTGTGAGATAGCCAAACGGAAGCTTCTATGAGAGATCATGATCC 1109
Db 1051 TTTTCCCGCGCTTCGATTCCTCGTAAACGGAAGATTTCAACCGAGTTCCTTTGTTTAA 1110
QY 1110 GGGAAAGGTATAAGATCCGGGAGAGAGTTGAGTGATTTACGTCACAGAGGAGA 1169
Db 1111 AGGAGAGCTTTACCGGATAAATTTGTCGCTTTATTACGCTGGGAATGCTAGTATGC 1170
QY 1170 TAAAGGAGTGAAGTCTCTTTGAGAGGTCACITCCAAAGAGAGAAATCCGAGCGCAAAAT 1229
Db 1171 TACTAATGGTAATCTCTGTATGACCGNACTTTGATCCGAGAGAAATGAAGGGAGAT 1230
QY 1230 GGTGATTTGATCCGCGAGTCAATGGAAGATCGGAGAAAGGAGACGTTAAAGAAC 1289
Db 1231 TGTGATGTGTATGAGGAATTAATGCTAGAGTTTCAAGAAAGTGATGTGTTAAAGCAGC 1290
QY 1290 TGGAGGATTCGATGATCTTAGCCATACAGATCAACCAAGAGAGATCTATTGA 1349
Db 1291 AGGTGGAGTTGGAATGATTTCTGGCTAACTCGCGCAATGGTGAAGAGCTTGTTCGGA 1350
QY 1350 CGTTCACTCTTACAGCTACATGATTTGTTTACACTGAGTCAGTCTCTTCTGAAGGCTTA 1409
Db 1351 TGCTCAATTTGTACCGGCAACACCGTTGGTGAAGAACCGGTGATATATCCGGCATTA 1410
QY 1410 TGTTAATCCACGCTGAACCAAGGCGCGGATAAATTTTGGTGGTACGGTGAATGGGAG 1469
Db 1411 CGTGACTACTGATCCTAATCCACCGCTTCCATTTCAATCTTAGAACAGCTCGTCGCTG 1470
QY 1470 GTCAGAGCAACGAGGTGGCTCAGTTTTCAGTTCGAGGACCGAGTTTACCAATCCTTC 1529
Db 1471 TAAACCATCTCCGTTGTCGACGCTTTAGCTCACTGCGGACCGGAATTCGATTACACGGA 1530
QY 1530 GATATAAACCGGATATGATGTCGCGGAGTCAATATCATTCCGCTTGGCTTCAAAA 1589
Db 1531 TATTTCTAAACCGGATCTGATGCTCTCTGGAGTAAACATCTCCCGCGTGGACCGGTGC 1590
QY 1590 TCTAGACCAACCGGACTTCCTTATGATTCAGGAAGAGTTAACTTCACTGTAAATGTCCAGG 1649
Db 1591 TGCTGACCAACCGGACTCGCTTCGATCTTCGCGGCTGGAGTTCAATATCACTCTCGG 1650
QY 1650 AACTTCAATGTCTTGTCCACATGTTAGCGGAATCACTCTTATCCGCTTCGATACCC 1709
Db 1651 GACGTCGATGCTTCCCTCAGTTAGTGGTTAGCGCGCTTCTCAAGTCGGTGCATCC 1710
QY 1710 GAACTGGTCTCAGCTGCAATCAATCCGCAATTCGATGATGACAAACAGCGGATTTGTACGATCG 1769

Db 1711 TGAATGTAGCCGCGGGCGATTAGATCGCCCTTATGACACCGCTTACAAAACCTACAA 1770
QY 1770 TCAACGGAAAGCGATA-----AAGGATGTAACAAACACCGCGGTGTGTTGCGATTGG 1823
Db 1771 AGACGCTAAACGTTTACTCGACATCGCACAGGGAAGCCTTCGACCGCGTTTCGATCACGG 1830
QY 1824 AGCAGGCGATGTGAATCCGCAAAAGGCGATAAACCCGGGATTTGTTTACACATTCACACC 1883
Db 1831 TGCAGGACACGCTGTCCACCAACATGCGCACTAATCCAGGACTCATCTACGATCTAAACGAC 1890
QY 1864 AGTGATATACATAACTTACCTCTGACATCTTGGATTACAAAGATCAGATATT 1935
Db 1891 GGAGGATTAATAGCTTCCCTCTCGCATTTGAATTACACATCGCGCGAAATT 1942
RESULT 12
ATARA12
LOCUS 2423 bp mRNA linear PLN 14-MAR-1996
DEFINITION A.thaliana mRNA for subtilisin-like protein.
ACCESSION X85974
VERSION X85974.1 GI:757533
KEYWORDS ara12 gene; subtilisin-like protease.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1
AUTHORS Ribeiro,A., Akkermans,A.D., van Kammen,A., Bisseling,T. and Pawlowski,K.
TITLE A module-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development
JOURNAL Plant Cell 7 (6), 785-794 (1995)
MEDLINE 95375542
PUBMED 7647567
REFERENCE 2 (bases 1 to 2423)
AUTHORS Ribeiro,A.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-1995) A. Ribeiro, Dept. of Molecular Biology,
Agricultural University Wageningen, Dreijenlaan 3, NL- 6703 HA
Wageningen, NETHERLANDS
COMMENT Related sequence: T04180.
FEATURES
Location/Qualifiers
1..2423
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="lambda-PRL2 (T Newman)"
/clone_lib="CDNA library derived from 4 pools of mRNA, (1) 7 days germinated, etiolated plants; (2) tissue culture grown roots; (3) staged plants half with 24 hours light cycle, half on 16 hours light, 8 hour dark rosettes; (4) as (3) but aerial tissue (stems, flowers and siliques)"
1..2423
/gene="ara12"
<1..2242
/gene="ara12"
/codon_start=2
/product="subtilisin-like protease"
/protein_id="CAA5963.1"
/db_xref="GI:757534"
/db_xref="SPTREMBL:Q39007"
/translation="AFFLLLCIGFCHVSSSSSDQGTIVHMAKSTPSSFDLHNSWYD
SSLRSISDAELLYTYENAIGFSTRLTOEADSLMTOPGVISVLPERHYELHTTRTP
LPLGLDEHADIPEAGSYSDVVGVLDTGVWPESKYSDEFGFIPSSWKGGCAGT
NFTASCKMLIGARFFARGVSTMGPIDESKESPRDDDGHTHTSTAGSVVEG
ASLLYAGSTAFGLHALAVKVCWLGCGFSSDILAIADKAIDNNVNLVSLGGMS
DYRDRGVAIGAFAAMERGLVSCSAGNAPSSSLSNVAPWITTVGAGTLDRDFPALLA
IILGNKFTVGLFKGALPKLLPFI VAGNASNATGNLCMTGLTIPEKVKGIWVC
DRGINARVQKGVKAGGVGMILANTAANGELVADAHLLPATTVGEKAGDIIRHY

Db 1159 CATTCAAAGGAGTGTCTCTTAAACAGGAGAGACAGTTTTCCTTAAGATAAACAGTATC 1218
Qy 1145 AGGTGATTATCGT-----CACAGGAGGAGATAAGCAAGTGTCTGTTTGGAGAG 1195
Db 1219 CTCTGGTTTACTTAGGAAGGATGCAAGTAGTCTGTGATCAACCTGCTGTCTGTAGATG 1278
Qy 1196 GGTCACTTCCAAGAGAGAAATCGAGGCAAAATGGTGAATTTGTATCGCGGAGTCAATG 1255
Db 1279 GAGCTTTGGATCGCGGCATGTAGCGGAAAGATCGTGATCGACCGCGGTGTACTC 1338
Qy 1256 GAAGATCGGAGAAAGGAGAGCGGTTAAAGAGCTGGAGGAGTGCAGATGATCTTAGCCA 1315
Db 1339 CACGTGCAAAAGGGTCAGGTTGTGAAAGAGAGCTGTGGAATGGGATGGTTTTAACTA 1398
Qy 1316 ATACAGAGATCAACCAAGAGAGATCTATTGAGCTTCTCTTACCAAGCTACATTTGA 1375
Db 1399 ACATGCAAAATGGTGAAGAGCTTGTTCAGATTTCTCATATGCTTCCAGCTTGTGCAG 1458
Qy 1376 TTGTTTACACTGAGTCAGTCTCTGAAAGGCTTATGTTAATGCAAGCGTGAACCAAGG 1435
Db 1459 TTGAGAGAGAAAGGTAACATAATCAACAGTACGGGATGACGAGTAAAGAGGACAG 1518
Qy 1436 CGCGATATTTTTGGTGTGACGTTGATGGAGGTCAAGAGCACCGGAGTGGCTCAGT 1495
Db 1519 CGAGTTTAGAGATCTTTGGAACAAGAAATGGTATCAAAACCTTCACCAAGTTGTAGACGCT 1578
Qy 1496 TTTGAGTCGAGGACCGAGTTTACCAATCTTCGATACTAAACCGATATGATGCTC 1555
Db 1579 TCTCTTCAAGAGGACCAAAATTTCTGCTTTGGAGATCTTGAACCAAGATTTGTGGCTC 1638
Qy 1556 CGGAGTCAATATCATTTGCGCTTGGCTCATAATCTAGGACCAACCGGACTTCTTTATG 1615
Db 1639 CAGGATGAATATTTTGCAGCTTGGACTGGAGACATGGCACCATCGAGTTTATCATCTG 1698
Qy 1616 ATCAAGAGAGTTAACTTCACTGTAATGTCAGGAATCAATGCTTTGTCACATGTTA 1675
Db 1699 ATCAAGAGGAGTTAAAGTTCAATATCTGTCGAACTTCAATGCTATGCTCCTCATGTA 1758
Qy 1676 CGGNAATCACTGCTTATCGGCTGCTGATACCGGACTGCTCCAGCTGCAATCAAT 1735
Db 1759 GTGGTGTAGCTGCTTTGATCAAAATCAAGGCATCCAGATGGAGTCTTCAGCAATCAAT 1818
Qy 1736 CGGATTCATGACACACGAGATTTGACGATCTGCAAGGAAAGCATTAAGATGGTA 1795
Db 1819 CAGCTCTCATGACAACTGCTTATGTTCTATGACACATGTTTAAAGCTTCTTAAGATCAT 1878
Qy 1796 ACAAACCGACCGT-----GTGTTGCGATGGAGCAGGCGATGTGAATCCGCAAAAG 1849
Db 1879 CAGGAGCAGCTCCTTCATCGCCTTATGATCAGCGTGCAGGACATATAGATCCTTTAAGAG 1938
Qy 1850 CGATTAACCGCGGATGTTTACACATTCACAGTGGATGATACATTAACCTTACTCTGCA 1909
Db 1939 CTACAGATCCTGGTTGGTCTACGACATTTGACCTCAAGATATTTTGAATTCCTCTGCA 1998
Qy 1910 CTCCTTGGATTCACAGATCAGATATTTTGAACATCACTCATAGAACGTCGAGTGCATG 1969
Db 1999 CTCAGATTTAGTCCATCAGCTTAGGTATTCACAAACATTCACAAAGACCTGCA 2058
Qy 1970 GAATATTCGGAAGAAACCGCGGTTTTAGTCTCAATTTACCCGTGATAGCGGTGATTTTCA 2029
Db 2059 AACACACTCTTGCCAGAAATCCGGAAACTTGAACCTACCCGGCGATATCAGCTTTGTTC 2118
Qy 2030 AACGTGGCA-----AGACTAGGAGATGATCAAGCGGTGCTCAATAGCTTGGAGTC 2083
Db 2119 CAGAGAACACATGTTTAAAGCTATGACACTTAGAAGAACAGTGAACCAATGTTGGTCTC 2178
Qy 2084 CTAATCTGATATCTACAGTGAATGTCAGGCTCCAGGGGATCAAAAGTTATTGCTCAATC 2143
Db 2179 ACATTTCAAGTACAAAGTTTCTGCTCCGATTCAAAGGCGCATCCGTAAGTCTCCAGC 2238
Qy 2144 CTAAGAGACTTGTGTTCAACACACGTTGATCAGAGCTGAGCTATAGAGT 2192
Db 2239 CCAAAACACTCACTTCACTTCGAGACCGAGAGCTTTCCTACAGCT 2287

RESULT 14

AB010074
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AB010074 86212 bp DNA linear PLN 27-DEC-2000
Arabidopsis thaliana genomic DNA, chromosome 5, Pl clone:MIO24.
AB010074 BA000015
AB010074.2 GI:10177866

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsels.

1 (sites)
Sato,S., Kaneko,T., Kotani,H., Nakamura,Y., Asamizu,E., Miyajima,N.
and Tabata,S.

Structural analysis of Arabidopsis thaliana chromosome 5. IV.
Sequence features of the regions of 1,456,315 bp covered by
nineteen physically assigned Pl and TAC clones

DNA Res. 5 (1), 41-54 (1998)

98290546

9628582

2 (bases 1 to 86212)

Nakamura,Y.

Direct Submission

Submitted (07-JAN-1998) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamura@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)

On Sep 15, 2000 this sequence version replaced gi:2760170.

Address for correspondence: kaos@kazusa.or.jp

For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd/graph.cgi?c=MIO24

Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.

The software programs used to predict genes include: Grail

(Informatics Group, Oak Ridge National Laboratory,

http://compbio.ornl.gov/Grail-1.3/).

GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and

splicePredictor (Volker Brendel, Stanford University,

http://grmbln1.zool.iastate.edu/cgi-bin/sp.cgi).

Genes encoding tRNAs are predicted by tRNAscan-SE

(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).

This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.

The 5' clone is K10P11 and the 3' clone is MJM18.

Location/Qualifiers

1. 86212

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia"

/db_xref="taxon:3702"

/chromosome="5"

/clone="MIO24"

/clone_lib="Mitsui Pl"

1494. 2471

/note="gene id:MIO24.19

unknown protein"

/codon_start=1

/evidence=not experimental

/protein_id="BAB11237.1"

/db_xref="GI:10177867"

/translation="MSDMEPKETKPPRKQLRPPSPVFIWEERPGFKKNWPSLATF
VSPSPPLPPPTPVKLVTSVPPFRWEETPGKELPASNDPQLPPEPLETATETPLPP
PVPPVKQTVSVPFDEETPGQPCFVDTSPPELLDQPLPFPYMGKGVESVDSIFDD
ASSDFSVPSSLATNRVTSIGAVAVDFDNLNTVTSMTSPAYSDSDTSYMT
GASLVGASFLFKLPRLLPSEKVAASVDSVQVSTHPLHEVLTITDNMSIGFPV
RTPQTGLGELIMMSRRRSYMRRAVENMRKQNPVTVRLSNIFSHSLPKPYAP"

FEATURES
source

CDS

Db 24497 CTGTGATACCTGAGACAAGATACAGGCTTCACACTACAAGAAATCCAAAGTTTCTTTGGGT 24556
Qy 365 TCACAGGTTTGGAACTCCGGTGTATGTCFAAATCTCGGTTTGTGTCRAAGCACAAATTA 424
Db 24557 TAGAAAACAAAGAAAGTGAGAGAGTTTGGGCTCAGAGAGTCCCGACCATGATGTGGTAG 24616
Qy 425 TCGGCGTGTGTGATACCTGGAAGTTGGCTGAAAGTCTCTAGCTTTGACGATACCGGAATGC 484
Db 24617 TTGGTGTTTAGACATGGTATCTGGCTGAGAGTGAAGCTTCAACGATACAGGTATGT 24676
Qy 485 CTTTCGATTCACCGAAATGGAAGAGGATTTGCCAAGAGAGAGAAATTTTCAGTTCTTCCA 544
Db 24677 CCCCTGTTCTGCTACTTGGAGAGGAGCTTGTGAAGTCTGGAAGAGATTTCTTGAACGTA 24736
Qy 545 GCTGTAAACCGAAGCTAATCCGTCCTAGATCTTTCATCAGAGACACCGGTGTCGTAAT 604
Db 24737 ACTGCAATGAAGAGATCGTTGGTCTAGAGTTTCTTATAGAGG-----CTATGAAG 24787
Qy 605 CACAGAGGAATCACCAAAATCGCTCGTGAATACATTTCCGCAAGAGATTCACCGGAC 664
Db 24788 CTGCAACGGGAAGATCGATGAAGAGCTTGAAATTAAGTCCAGAGAGACAGATGGTC 24847
Qy 665 ACGGACTCACACCGCTCAACAGTTGGTGGATCTCTGTTTCGATGGCGAATGTTCTTG 724
Db 24848 ACGGACACACACTGCAGCTACTGTAGCTGGCTCACTGTAAAGGAGCTAATCTTTTG 24907
Qy 725 GCAATGAGCTGGTGGCTGCTGGGATGGCTCTGGAGCTCACATGTGAGTCTATAAG 784
Db 24908 GATTTGCTTATGGGACAGCTCGAGGAGTGGCTCAAAAGGCTAGAGTGTGCTTATATAAG 24967
Qy 785 TCTGTTGGTCAATGGTTGTAGAGCTCTCACATTTAGCAGCTATAGATGTAGCGATTC 844
Db 24968 TCTGTTGGTGGAGGTTTTCAGTTCAACATTTTTCGGCTGTTGATCAAGCTGTTG 25027
Qy 845 AAGATAAGTCGATGTTCTTTGCTTTCCCTTTCCGCTTTCCCTATTCTTTGATGATG 904
Db 25028 CTGATGAGTTCAGTACTCTATATCATTAAGTGGTGGGCTCTACTTATCTTAGAG 25087
Qy 905 ACACAAATGGCAATGGAACATTCGAGCCATGGAACCGCGTATATCTGTAATCTGTGAG 964
Db 25088 ACAGTTGTCTATAGCAACATTTGGAGCAATGGAGATGGAGTTTTCGTTGCTGTTCTG 25147
Qy 965 CTGTAACACGGTCCATCGAAGCTCTGTCGAACACAGCTCTCTGGGCTCTCAACCA 1024
Db 25148 CCGTAATGAGAGTCTGATCCGATAGTCTCACTAATGTTTCTCCATGGATCACACAG 25207
Qy 1025 TTGGCGAGGACACCTTGTATCGAAGATTTCCGCTGTGCTCAGATTAGCCAAACGGAAGC 1084
Db 25208 TTGTTGCAAGTACTATGATAGATTTTCCAGCAACAGTGAAGATAGGAATATAGAA 25267
Qy 1085 TTCTCTATGAGAGTCAATGATCCGGGAAAGGATATAAGAAATGCCGGGAGAGAGTTG 1144
Db 25268 CATTCAAAGGAGTGTCACTTTTACAAAGGCAACAGATTTTGCCTAAGATAAAGACAGTATC 25327
Qy 1145 AGGTGATTTAGT-----CACAGAGAGATAAGGAAGTGTCTGTTGTAGAG 1195
Db 25328 CTCTGTTTACTTAGGAAGGAATCAAGTAGTCTCTGATCCAACTCTGCTCTAGATG 25387
Qy 1196 GGTCACTTCCAGAGAGAAATCCGAGCAAAATGGTGTATGATCGCGGAGTCAATG 1255
Db 25388 GAGCTTTGGATCGGGCCATGTAGCGGAAAGATCTGTAATGCGACCGGGTGTACTC 25447
Qy 1256 GAAGATCGGAGAAAGGAGCGGTTAAAGAGCTGGAGGAGTTGCAATGATCTTAGCCA 1315
Db 25448 CACGTGTCAAAAGGTCAGTTGTGAAAGAGCTGTGGAAATGGGATGTTTAACTA 25507
Qy 1316 ATACAGAGATCAACCAAGAGAAATCTTATGAGCTTCACTCTTACAGCTACATGA 1375
Db 25508 ACCTGCAACAAATGGTGAAGAGCTTGTGAGATTTCTCATATGCTTCCAGCTGTTGAG 25567
Qy 1376 TTGGTTACACTGAGTCACTCTCTGAGGCTTATGTTAATGCCAGCTGAAACCAAGG 1435
Db 25568 TTGGAGAGAAAGGTAACCTAATCAACAGTAGCGGATGACGAGTAAAGGCGAG 25627

Qy 1436 CGCGGATAAATTTTGTGGTACGTTGATGGGAGGTACGAGCACCGGAGGTGGTTCAGT 1495
Db 25628 CGAGTTTATAGAGATTTCTTGGAAACAAGAAATGGTATCAAACTTCCACAGATTTGTAGACGCT 25687
Qy 1496 TTTGAGTCTCAGAGACCGAGTTTACCAATCTTCGTACTATAAACCGGATATGTTGCTC 1555
Db 25688 TCTCTTCAAGAGGACCAAAATTTTCTGTCTTTGGAGATCTTTGAACACAGACTTTGTGGCTC 25747
Qy 1556 CGGAGTCAATATCATTTGGGCTTGGCTCAAAATCTTAGACCAACCGGACTTCCCTTATG 1615
Db 25748 CAGAGTGAATATTTCTTGCAGCTTGGACTGGAGACATGGACCATCGAGTTTATCATCTG 25807
Qy 1616 ATTCAAGAAGAGTTTAACTTCACTGTATGTGAGAACTTCAATGTCTTTGTCACATGTTA 1675
Db 25808 ATCCAAGAGGTTAAGTTCAATATCTGCTGGAACCTTCAATGTCTATGCTCTCATGTAA 25867
Qy 1676 GCGAAATCACTGCTCTTATCCGGTCTGCATACCGGAATCTGCTCCAGCTGCAATCAAT 1735
Db 25868 GTGGTGTAGCTGCTTTGATCAAAATCAAGGATCCAGATTGGAGTCTCTGACAGCAATCAAT 25927
Qy 1736 CCGCAATTGATCACACAGCGGATTTGTACGATCGTCAAGGAAAGCGATAAAGATGTA 1795
Db 25928 CAGCTCTCATGACAACTGCTTATGTTTATGACAACTGTTTAAAGCTTTCAGGATGCAT 25987
Qy 1796 ACAAACAGCGCGT-----GTGTTTGGATTTGAGCAGGCGCATGTGAATCCGCAAGG 1849
Db 25988 CAGAGCAGCTCTTCTCATCGCTTATGATCAOGTGCAGGACATATAGATCCCTTAAAGAG 26047
Qy 1850 CGATAAACCCGGGATTTGTTTACAACTCAACCAAGTGGATATACAACTTACCTCTGCA 1909
Db 26048 CTACAGATCTCTGTTTGGTCTACGACATTTGACCTCAAGAGTATTTTGAATTCCTCTGCA 26107
Qy 1910 CTCCTGATTCACAGATCAGATATTTTAGCAATCACTATAAGAACGTGAGCTGCAATG 1969
Db 26108 CTCAGATTTAAGTCCATCAGCTTAAGTATTCACAAACATTCAAACAGAACCTTGCA 26167
Qy 1970 GAATATTCCGGAACAAACCCGGGTTTGTCTCAATTTACCTCCGATCGATAGCTGATTTTCA 2029
Db 26168 AACACACTTCTGCGAAGATCCGGGAAACTTGAACCTACCGCGGATATCAGTTGTTCC 26227
Qy 2030 AACCTGGCA-----AGACTACGGAGATGATCAAGCGGTGTCACAACTGTTGGAGTC 2083
Db 26228 CAGAGAACACACATGTTTAAAGCTATGACACTTAGAAGAACAGTACCAATTTGTCCTC 26287
Qy 2084 CTAACTCGATATCTCAGTGAATGTCAAGGCTCAGAGGGATCAAGTATTTGTCATC 2143
Db 26288 ACATTTCAAGTCAAGGTTTCTGCTCGCAATTTCAAAGGCGCATCCGTAACCTGTCAGC 26347
Qy 2144 CTAAAGAGCTTGTGTTCAAAACAGCTGGATCAGACGCTGAGCTATAGAT 2192
Db 26348 CCAAAACACTCAACTTCACTTCGGAAGCACCAGAGCTTTCCTACAGGT 26396

RESULT 15

AF360129

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AF360129 2525 bp mRNA linear PLN 18-SEP-2002
Arabidopsis thaliana putative subtilisin serine protease
(At3g14240) mRNA, complete cds.

AF360129.1 GI:13430433

FLI_CDNA.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 2525)

Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,

Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,

Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,

Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,

Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,

Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
Davis R.W., Ecker J.R. and Theologis A.
Arabidopsis Full Length cDNA Clones
Unpublished
2. (bases 1 to 2525)

Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B.,
Lin J., Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T.,
Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
Davis R.W., Ecker J.R. and Theologis A.

Direct Submission
Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA

RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-length cDNA'): Seki M., Narusaka M., Ishida J.,
Satou M., Kamiya A., Sakurai T., Carninci P., Kawai J.,
Hayashizaki Y. and Shinozaki K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada K., Liu S.X.,
Sakano H., Pham P.K., Banh J., Chung M.K., Goldsmith A.D.,
Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L., Chen H.,
Cheuk R., Jones T., Karlin-Neumann G., Kim C., Lam B., Lin J.,
Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A.,
Davis R.W., Ecker J.R. and Theologis A.

Yamada K. (SSP/PGEC) and Seki M. (RIKEN GSC) contributed equally to
this work. Shinozaki K. (RIKEN GSC) and Theologis A. (SSP/PGEC)
contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to Genbank.

FEATURES

source

Location/Qualifiers

1..2525

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="3"

/clone="RAFL09-15-B12 (R09108)"

/note="This clone is in a modified plusuescript vector
(FLC-1) as a BamHI/XhoI insert."

ecotype: Columbia

1..2525

/gene="At3g14240"

1..38

/gene="At3g14240"

39..2366

/gene="At3g14240"

/note="subtilisin proteinase-like"

/codon_start=1

/evidence=experimental

/product="putative subtilisin serine protease"

/protein_id="AAK25839.1"

/db_xref="GI:13430434"

/translation="MARFSYFFLLTLSSSSSSSSNSLTYIVVHDHEAKPSIFPTH
FHWTSLSASITSPSPSIITHTVDFHGFSAFLTSQASQLDHPHVISVPEQVRHL
HTTSPFLGLRSTDKAGLLEESDFGLVIGDITGVWPFSPFDDRGGLGVPFIKWK
QCQIASQDFPESACNRKLVGAFPCGGVYEAATNGKNETTIFRPSRDSGDHGTASIS
AGRVFPASTGLYGAHGAAGVAPKARLAAYKVCWNSCYDSDIILAAPDTAVADGVDI
SLSGVGVVFPYLLAALGAFGADRGIFVYSAGNGGPGALFTVNPWMTTVGAGT
IDRFPANVLGNKMGISVSGVPLDPMYPLVIGSLGGDGYSSSLCEGL
DNLVKGKVLICRGINSRATKBEIVRKNGLGMIANGVDFDGLVADCHVLPATSV
GASGDEIRRYISSSSKSRSSKHTATVFKGRLGIRPAPVVASFASGPNPPEI
LKPVPVIAQNLILAAAPDRIGSPGTVSDNRTEPILSGTSMACPHVGLAALLKAAH
PDWSPVATRSALITATVDNSGPMWDESTGNTSSVMDYSGSHVHTKAMDPLVVD
ITSVDYINFLCNSYTRNIVITERRQDCGARRAGHGNLNYPSFVVFPQVGSK
MSTHRTVTVNGSDSVYEIKIRPPGTIVTEPEKLSFRRVQGLKFLSVVRVKTTEV
KLSFGATVETGHTVMSDGRNVTSPLVVTLQQL"

52

/gene="At3g14240"

/note="compared to genomic sequence resulting in an amino
acid sequence difference"

/replace="t"

2367..2525

/gene="At3g14240"

misc_difference 2508

/gene="At3g14240"

/note="compared to genomic sequence"

/replace="a"

BASE COUNT 585 a 584 c 623 g 733 t

ORIGIN

Query Match

Best Local Similarity 18.2%; Score 424.4; DB 8; Length 2525;

Matches 1075; Conservative 0; Mismatches 916; Indels 51; Gaps 5;

QY 200 CTTCTTCTCGAGCTTCTTACTCTCTATGCTGCTCGATTGAAGAGATTGTGCTCAGTTGA 259

Db 211 CTCTCCCTCCATTATCCACATACGACACCGTTTTCATAGGCTTCTCCGTACACTCA 270

QY 260 CTGAATCAGAAGCCGAGATCTAGATATTCACCTGAAAGTTGTTGCAGTGAGACTGACC 319

Db 271 CTTCAAGAGCCTAGTCAACTCTTTAGACCATCTCTCATGTCTCTGTATCCCGGAAC 330

QY 320 ATGTTCTTCAGGTTCAAAACCACTTACTCTTACAGTTCTTGGGACTCCAGCGTTTGGAA 379

Db 331 AAGTCCGTCACCTTGCACACCACTCGTCCCTCGAGTTTCTTGGTCTTAGGTCCACCGACA 390

QY 380 ACTCGGCTATGTTCTAAATCTCGTTTGGTCAAGGCACCAATTATCGCGTGTCTGATA 439

Db 391 AAGCGGCTCTCTGAAGAGTCTGATTTCCGGTCCGATCTAGTTATCGAGTTATCGATA 450

QY 440 CTGGAGTTTGGCTGAAAGTCTCTAGCTTTGACGATACCGGAATGCTTCGATTCACCGGA 499

Db 451 CTGTGTTTGGCCCGAAGACCTAGCTTTGATGACCGTGGTCTTGGTCTGTTCCTCATTA 510

QY 500 AATGGAAGGATTTGCCAAGAGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGAAGC 559

Db 511 AATGGAAGGCCAATGTATCGCTTCCCAAGATTTTCCGAGTCTGTTGTAACCGTAAAC 570

QY 560 TAATCGGTCTAGATTTCTTCATCAGAGGACACCGTGTGCTTAATTCACAGAGGAATCAC 619

Db 571 TCGTCGGAGCTAGATTTCTTCGCGTGGGTA-----TGAAGCAACCAACCGGAAAA 621

QY 620 CAAACATGCTCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCG 679

Db 622 TGAATGAACGACTGAGTTTCCTCTCCGCTGACTCCGATGACATGGGACTCACACAG 681

QY 680 CCTCAACAGTTGGTGGATCCTCTGTTTCGATGGCGAATGTTCTTGGCAATGGAGCTGTG 739

Db 682 CTTCAATCTCCGCGCGCGTTACGTTTTTCCGGGCTCAACTCTCGGCTACGCTCAGCGG 741

QY 740 TGGCTCGTGGATGGCTCTCGAGCTCACATTGCGAGTCTATAAAGTCTGTGTTGTTCAATG 799

Db 742 TCGCTGCTGGGATGGCTCCAAAAGCTAGACTCCGCGCTTACAAAGTCTGTGTTGAACTCCG 801

QY 800 GTTGTATACAGCTCTGACATTTCTAGAGCTATAGATGTAGCGATTCAAGATAAAGTCGATG 859

Db 802 GTTGTATACAGCTCCGATATCTTTAGCCGCTTTCCGACACCGCTTCCGCGAGTGTGATG 861

QY 860 TTCTTTTCGCTTTCCCTTTGCGGTTTCCCTATTCCCTTTGTTATGATGACACAAATCGCATG 919

Db 862 TCATCTCTCTCCGTTGGAGCGCTGCTGTTTCTTATTACCTAGACGCTATCGCTATTG 921

QY 920 GAACATTCCGACCCATGGAACCGGTTATCTGTAATCTGTGCACTGGTAAACAACGGTTC 979

Db 922 GAGCTTTTGGAGCTATTGACAGAGCATATTGCTCTGCTTCCGCGCTTAACGGAGGTC 981

QY 980 CAAATCGAAGCTCTGTTGCAACAGCTCTCTTGGCTCTCAACATTGGCGCAGCACCGC 1039

Db 982 CCGGTGCTTTAAACGTTGACAAATGTTGCTCCGTGGATGACACAGCTCGAGAGTGAACAA 1041

QY 1040 TTGATCGAAGATTTCCGCTGTGGTTCAGATTAGCCAAACGGAAGAGCTTCTCTATGAGAGT 1099

gene

5' UTR

CDS

misc_difference 52

Db 1042 TCGATAGGATTTTCGGCCAAATGTGAACTCGGTAAACGGGAAGATGATTTCTGTGTGTTA 1101
Qy 1100 CATGTATCCGGGAAAGGTATAAA-----GAATCCGGAGAGAGGTGAGGTGA 1150
Db 1102 GTGTATACGGTGGACCGGGTCTGGATCCGGGTGGAATGTACCGCTTGTGTTACGGTGGTA 1161
Qy 1151 TTTACGTCAACAGGAGAGATAAAGGAAGTGAAGTTCGTGTTTGAGAGGGTCACCTTCCAAGAG 1210
Db 1162 GTTTACTAGCGCGGATGGTTACTCATCGTCTCTGTGCTTGAAGGCTCGTTGGATCGGA 1221
Qy 1211 AAGAAATCCGAGGCAAAATCGTGAATTTGATCCGGAGTCAATGGAAGATCGGGAAG 1270
Db 1222 ATTTAGTGAAGGAAGATCGTCTTGTGATAGAGGAATCAATCTAGAGCAACCAAG 1281
Qy 1271 GAGAGCGGTAAAGAGCTGGAGGAGTGCATGATCTTAGCCATACAGAGATCAACC 1330
Db 1282 GTGAGATCGTACGGAATAATGGAGCTTGGGGATGATATAGCAATGGTGTGTTGAGG 1341
Qy 1331 AAGAGAAGATCTATTGACGTTTCATCTTACCACTACATGATTTGGTTACACTGAGT 1390
Db 1342 GTGAAGTTTAGTAGCTGATGCGACGTTTACCGGGACATCTGTGTTGTTCTGGAG 1401
Qy 1391 CAGTCTTCTGAGGCTTATGTTAATGCCACGGTGAAA-----CCAA 1432
Db 1402 GAGATGAGATTCGTAGGTATATCTCGAATCATCCAAATCTCGTTCATCGAAACATCCAA 1461
Qy 1433 AGCGCGGATAATTTTGGTGGTACCGGTGATGGAGGTCAAGAGCACCGAGGTGGCTC 1492
Db 1462 CGGCTACGATGTTTTCAAAAGGTACCGACTTGGGATCCGACCTGCAACCGGTTGTTGCAT 1521
Qy 1493 AGTTTTTCAGGTCGAGGACCGAGTTTACCAATCTTTCGATCTAAACCGGATATGATG 1552
Db 1522 CTTTCTCTGCTGCGGCTTAAATCCAGAGACCGCGGAGATCTTAAACCGGATGTAATCG 1581
Qy 1553 CTCGGGAGTCAATCATTCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTTCCTT 1612
Db 1582 CACTGTGTTGAATTTCTAGTCTTGGCTGACCGATTGTCCTCATCTGGTGTACTT 1641
Qy 1613 ATGATCAAGAAGAGTTAACTTCACTGTAATGTAGGAACTTCAATGTCTTGTCCACATG 1672
Db 1642 CTGATAACCGGAGAACTGAGTTCAAATTTTATCAGGCACTTCGATGGGTGCGCCGACG 1701
Qy 1673 TTACGGAACTACTGCTCTTATCCGCTCTGCATACCGGACTGGTCTCCAGCTGCAATCA 1732
Db 1702 TGTCTGCTAGCTGCTTTCCTTAAGCGGCTCATCCGATGGAGTCCGGCGCGGATAA 1761
Qy 1733 AATCGGATTTGATCACAACAGCGGATTTGTACGATCGTCAAGGAAAGCATAAAGGATG 1792
Db 1762 GATCAGCATTTGATTAACCAACCGCTTACAGGTTGATTAACAGCGGTGAGCCGATGATGATG 1821
Qy 1793 -----GTACAAACCCAGCGGTGTGTTGCGATTGGAGAGGGCATGTGAATCCGCAA 1846
Db 1822 AGTCCACTGGCAATACATCTTCGGTTATGGATTATGGTTCGGGTCTATGTTTCAACCAACCA 1881
Qy 1847 AGCGGATAACCCGGGATTTGTTTAAACATTTAAACAGTGGATTACATTAACCTACTCT 1906
Db 1882 AAGCTATGGATCCGGGATTTAGTCTATGATATAACATCTTATGATTATATCAACTTCTGT 1941
Qy 1907 GCATCTTGGATTACAGATCAGATATTTTAGCAATCACTCATATAAGACGTTGAGTGA 1966
Db 1942 GTAAATTTAATACATAGAACCAACATTTGACAAATAACCCGTGCGCAAGCGGACTGTG 2001
Qy 1967 ATGGAATATTGCGGAAAAACCCGGGTTTTAGTCTCAATTAACCGTCTGATAGCCGTGATTT 2026
Db 2002 ACGGTGCAAGACGAGCGGACACGTCGGGAATTTAACTACCCGTCTGTTCTCTGTCGTGT 2061
Qy 2027 TCAA-----ACGTGGGAAGACTACGGAGATGATCAAGCGGTGTCACTAACGTTG 2077
Db 2062 TTCAGCAGTATGGAGAGAGTAAATGTGCACACATTTTATTAGGACTGTGACAAATGTAG 2121
Qy 2078 GGAGTCTTAACCTGATATCTACGTGAATGTCAAGGCTCCAGAGGGGATCAAGTTATTG 2137

Db 2122 GCGATTCCGATTTCGGTTTACGAGATTAGATTCAGGCGCTAGAGGACTACGGTGACGG 2181
Qy 2138 TCAATCCTAAAGAGACTTGTGTTCAACACACGTGGATCAGACGCTGAGCTATAGAGTATGGT 2197
Db 2182 TTGAGCCGGAGAACTATCGTTTCAAGCGGTGGGGCAGAACTGAGTTTGTGTTAGGG 2241
Qy 2198 TT 2199
Db 2242 TT 2243

Search completed: February 2, 2004, 23:28:56
Job time : 5720 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 3, 2004, 00:26:16 ; Search time 5004 Seconds
(without alignments)
6335.923 Million cell updates/sec

Title: US-09-806-767-2
Perfect score: 4018
Sequence: 1 MEKPFPLCHIFLLFCSSSS.....NSHNLQVRSPISVTLKTN 775

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgm2_1/USPto_epool/US09806767/runat_02022004_085958_17897/app_query.fasta_1.967
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09806767.ecgn_1_1_3296 @runat_02022004_085958_17897 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_by.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_io.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rnd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4018	100.0	78259	8	F20D22 Arabidops
2	2439	60.7	150503	8	AC122149 Cryza sat
3	1742	43.4	2266	6	AX506370 Sequence
4	1742	43.4	2266	8	AY142613 Arabidops
5	1742	43.4	2541	8	AY035090 Arabidops
6	1706.5	42.5	154863	2	OSJN00071 Arabidops
7	1700	42.3	83906	8	AB022220 Arabidops
8	1698	42.3	2525	8	AF360129 Arabidops
9	1697.5	42.2	121160	8	AC005970 Arabidops
10	1684	41.9	2638	8	AY084387 Arabidops
11	1679	41.8	2650	8	LESB2 Arabidops
12	1679	41.8	5197	8	LES6379 Arabidops
13	1666.5	41.5	2305	8	AY142612 Arabidops
14	1666.5	41.5	2618	8	AF360285 Arabidops
15	1666.5	41.5	2686	8	AF065639 Arabidops
16	1666.5	41.5	72698	8	AB007645 Arabidops
17	1662.5	41.4	2645	8	AF436834 Arabidops
18	1656	41.2	128017	8	AC084282 Arabidops
19	1653.5	41.2	2274	8	BT001082 Arabidops
20	1653.5	41.2	2520	8	AY091773 Arabidops
21	1642.5	40.9	2549	8	AY080831 Arabidops
22	1642.5	40.9	86212	8	AB010074 Arabidops
23	1640.5	40.8	2423	8	ATARA12 Arabidops
24	1611	40.1	2689	8	LESB1 Arabidops
25	1611	40.1	5089	8	LES6378 Arabidops
26	1600	39.8	2390	6	AX654166 Arabidops
27	1600	39.8	171099	2	AP003871 Arabidops
28	1593	39.6	145735	2	AC109601 Arabidops
29	1589	39.5	2295	6	AX506491 Arabidops
30	1589	39.5	2326	8	AY096357 Arabidops
31	1589	39.5	2665	8	AY074375 Arabidops
32	1589	39.5	80346	8	ATM4813 Arabidops
33	1589	39.5	103150	8	ATF1111 Arabidops
34	1589	39.5	195165	8	ATCHRIV82 Arabidops
35	1572.5	39.1	2847	8	LEAJS17 Arabidops
36	1565	38.9	199414	2	AC124955 Arabidops
37	1562	38.9	2465	8	BT002840 Arabidops
38	1554	38.7	113049	8	AC134239 Arabidops
39	1547.5	38.5	172769	2	AP005570 Arabidops
40	1542	38.4	2680	8	LEAJS172 Arabidops
41	1534.5	38.2	2407	8	LESUBT11 Arabidops
42	1534.5	38.2	3076	8	LES17276 Arabidops
43	1521	37.9	2883	8	LES6786 Arabidops
44	1513.5	37.7	2849	8	LEAJS173 Arabidops
45	1509.5	37.6	3173	8	LES17278 Arabidops

ALIGNMENTS

RESULT 1

```

F20D22      78259 bp      DNA      linear      PLN 20-MAY-1998
LOCUS       Arabidopsis thaliana chromosome 1 BAC F20D22 sequence, complete
DEFINITION
ACCESSION   AC002411
VERSION     AC002411.1  GI:2570223
KEYWORDS    HG
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
AUTHORS      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE    1 (bases 1 to 78259)
              Vysotskaia,V.S., Osborne,B.I., Schwartz,J.R., Toriumi,M., Kwan,A.,
              Yu,G., Oji.,O., Liu,S., Li,J., Hoang,L., Araujo,R., Au,M.,
              Brendel,V., Buehler,E., Conway,A.B., Conway,A.R., Dewar,K.,
              Feng,J., Kim,C., Kurtz,D., Li,Y., Palm,C.J., Shinn,P., Sun,H.,
              Davis,R.W., Ecker,J.R., Federspiel,N.A. and Theologis,A.
              Arabidopsis thaliana chromosome 1 BAC F20D22 complete sequence
              Unpublished (1998)
REFERENCE    2 (bases 1 to 78259)
              Theologis,A.
              Direct Submission
              Submitted (12-AUG-1997) Plant Gene Expression Center, 800 Buchanan
              Street, Albany, CA 94710, USA
              3 (bases 1 to 78259)
              Theologis,A.
              Direct Submission
              Submitted (28-OCT-1997) Plant Gene Expression Center, 800 Buchanan
              Street, Albany, CA 94710, USA
              4 (bases 1 to 78259)
              Theologis,A.
              Direct Submission
              Submitted (20-MAY-1998) Plant Gene Expression Center, 800 Buchanan
              St., Albany, CA 94710, USA
COMMENT      On Oct 28, 1997 this sequence version replaced gi:2323245.
              The sequence of BAC F20D22 from Arabidopsis thaliana chromosome 1.
              This sequence is of BAC F20D22 from Arabidopsis thaliana chromosome
              1. The sequence does not represent the sequence of the entire
              insert of this clone. It is shorter by 38223 bp because we sequence
              overlapping section only once. In order to facilitate the joining
              of overlapping clones in the future for creation of larger contigs,
              we provide a small overlap (200 bp) between overlapping submitted
              clones. The 5' end of this sequence overlaps by 200 bp the 3' end
              of the sequence of the BAC F19P19.
FEATURES             Location/Qualifiers
     source           1..78259
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F20D22"
                     3394..6024
                     /gene="F20D22.1"
                     join(3394..4161,5206..6024)
                     /gene="F20D22.1"
                     /note="Strong similarity to beta-keto-CoA synthase
                     gb|U37088 from Simmondsia chinensis."
                     /codon_start=1
                     /evidence=not experimental
                     /protein_id="AAC16740.1"
                     /db_xref="GI:3142289"
                     /translation="MNEHITQSDHMNTIHTVKKLPNFLLSVLRKYKVLGYHLISN
                     AVYILFLVGLLAATSSFTLDTLLYNHLKPHFLSSFTLFAALIFLITLYFTTRP
                     RYFLLPFCYKQDSSILCTRETFMDRSQRVGFTEDNLAFOQKILERSGLGKQYTFP
                     EALIRVPNFCMSEARKEATVWFGADVLEKTVGNPKDGIILVNCSLFNPTSLIS
                     AMIYVKYKLRVLSLYNMGWCSAGLISDLAKQLQVQPSYALVSTNTLAWY
                     LGNDRSMLLNCILFMCGGAVALLSNRRSDRCYQLIHTVRTHKSDDNAPCYVR
                     EDNDNQGVLSYKSNLMAAGEALKNTITLGLVLSMEQLFPATLYARVKVFNK
                     WEIYLAYSAGRKIRKGRDRTWQIAAGRAVLDEIEKNLDLSSEHMEQSRMTLRNFGNTSSSL
                     HEPFVVPRTSPVTSSESR"

```

gene /translations="MAEKKGATNGGAEKSLKEKGNPEFFKAGNFKAAALVTOAILK
DPSNATLYNRAAASLIVKLSKALADAETITIKLPQWEKVKSIHNDGFRKGCGLVLEA
MERIEDALAFENALQNFQSTSEVSRKIRLQOLQKQKQRAQELNENLNDVMAHGLE
SPFSESENGTQECNMFSPFIVETMETAVKSWHETSVDTRVFLIDKERTQDKY
APAVNDIKAFQSDPTNSNCFPIRQYABESFKAACLVTPKSSISYPOVWKGVSGRKW
KLGNPDIFQFSPSLRKVWFIFSSSKGQTLCRNRKARRALFCCSMILYIQLIK
TKIKSSCECVLLKSTFLYLI"
complement (19889..21254)
/gene="F20D22.5"
CDS complement (join(19889..20488,20589..21254))
/gene="F20D22.5"
/notes="Contains similarity to myosin IB heavy chain
gb|X70400 from Gallus gallus."
/codon_start=1
/evidence=not experimental
/protein_id="AAC16744.1"
/db_xref="GI:3142293"

gene /translations="MENPRLMASEYFSESRVCVWNGPVIIVGAGPSGLATAACILHDO
GVPPVVERSDCTASLNQKRTYDRLLKHLPKFKCQLPKMPFPDHPPEYPTKRFQIDYL
ESVANRFDIKPEFNKSVESARFDETSGLMRVTTSDGEMEYICRWLVATGENAERV
VPINGLMTFEDGEVTHACBYKSGEGRKRVLVVCGNSGMEVSLDLANHNAITSMV
VRSGVHLPREIMGKTFGTISVMKWLPLVLDKLLLSLWLVLSISNLYGLKRPDI
GMELASMTGKTPLVDIGALEKTKSGDVEIVPAIKOFSRHVELVDGOKLIDDAVLA
TGYRSNVPSWLQSESRFSKNGPEKSPFPNANKSKSLYAAGFRKGLAGASVDVAVNIA
QDIGNVRETKKQKMRNVGHRRCISVA"
complement (26175..28454)
/gene="F20D22.6"
CDS complement (join(26175..26477,26580..26840,26941..27159,
27271..27498,27776..27878,27985..28077,28197..28258,
28346..28454))
/gene="F20D22.6"
/notes="Strong similarity to initiation factor eIF-2,
gb|U37354 from S. pombe. ESTs gb|T41979, gb|N37284 and
gb|N37529 come from this gene."
/codon_start=1
/evidence=not experimental
/protein_id="AAC16745.1"
/db_xref="GI:3142294"

gene /translations="MSRNLGLAEOBODLKLDTVLHPLSPVISROATINIGTIGHVAH
GKSTVVKAGSVGTQVFKNELENIITIKLYANAKIYKCEDEKCPKMYKAYGSGKE
DTNCDVGFSESNKMLKHLRVSFQDPCGHIDIMATVNLNGAAIMDGLALLIANETCPQ
PTQSEHLAAVEINQOLHIIILQKIDLIQENVAINQHEAIQKFIINTVADAAPIVFVS
AQLKYNIDVCEIVKPIPIERNFVSPNMIIVRSFDVNFKEVDEIKGVAAGCSI
LRGLVARNQIIEIRPGIVTKDERGNSKCTPIYSRIISLVAEQNELQFAPVGLIGVGT
TMDPTLTRADRLGVGLGEISLPDVELEYNFLLRELLGVRTKSGEKQKSVKLT
KGIILMLNIGSNSTGAKVGVKVDLAKLQITAPVCTSKGEKVALSERVEKHRLIGWG
QIQAGTTTIEVPSPP"
complement (22906..38907)
/gene="F20D22.7"
CDS complement (join(22906..30073,30156..30212,30303..30359,
30434..30530,30700..30770,30848..30976,31135..31326,
31413..31565,31882..32052,32135..32185,32292..32336,
32427..32541,32629..32768,32889..33101,33207..33305,
33456..33575,33746..33951,34044..34221,34315..34375,
34464..34573,34672..34803,34895..35062,35184..35310,
35464..35573,35672..35803,35895..36062,36184..36310,
36464..36573,36672..36803,36895..37062,37184..37310,
37464..37573,37672..37803,37895..38062,38184..38310,
38464..38573,38672..38803,38895..39062,39184..39310,
39464..39573,39672..39803,39895..40062,40184..40310,
40464..40573,40672..40803,40895..41062,41184..41310,
41464..41573,41672..41803,41895..42062,42184..42310,
42464..42573,42672..42803,42895..43062,43184..43310,
43464..43573,43672..43803,43895..44062,44184..44310,
44464..44573,44672..44803,44895..45062,45184..45310,
45464..45573,45672..45803,45895..46062,46184..46310,
46464..46573,46672..46803,46895..47062,47184..47310,
47464..47573,47672..47803,47895..48062,48184..48310,
48464..48573,48672..48803,48895..49062,49184..49310,
49464..49573,49672..49803,49895..50062,50184..50310,
50464..50573,50672..50803,50895..51062,51184..51310,
51464..51573,51672..51803,51895..52062,52184..52310,
52464..52573,52672..52803,52895..53062,53184..53310,
53464..53573,53672..53803,53895..54062,54184..54310,
54464..54573,54672..54803,54895..55062,55184..55310,
55464..55573,55672..55803,55895..56062,56184..56310,
56464..56573,56672..56803,56895..57062,57184..57310,
57464..57573,57672..57803,57895..58062,58184..58310,
58464..58573,58672..58803,58895..59062,59184..59310,
59464..59573,59672..59803,59895..60062,60184..60310,
60464..60573,60672..60803,60895..61062,61184..61310,
61464..61573,61672..61803,61895..62062,62184..62310,
62464..62573,62672..62803,62895..63062,63184..63310,
63464..63573,63672..63803,63895..64062,64184..64310,
64464..64573,64672..64803,64895..65062,65184..65310,
65464..65573,65672..65803,65895..66062,66184..66310,
66464..66573,66672..66803,66895..67062,67184..67310,
67464..67573,67672..67803,67895..68062,68184..68310,
68464..68573,68672..68803,68895..69062,69184..69310,
69464..69573,69672..69803,69895..70062,70184..70310,
70464..70573,70672..70803,70895..71062,71184..71310,
71464..71573,71672..71803,71895..72062,72184..72310,
72464..72573,72672..72803,72895..73062,73184..73310,
73464..73573,73672..73803,73895..74062,74184..74310,
74464..74573,74672..74803,74895..75062,75184..75310,
75464..75573,75672..75803,75895..76062,76184..76310,
76464..76573,76672..76803,76895..77062,77184..77310,
77464..77573,77672..77803,77895..78062,78184..78310,
78464..78573,78672..78803,78895..79062,79184..79310,
79464..79573,79672..79803,79895..80062,80184..80310,
80464..80573,80672..80803,80895..81062,81184..81310,
81464..81573,81672..81803,81895..82062,82184..82310,
82464..82573,82672..82803,82895..83062,83184..83310,
83464..83573,83672..83803,83895..84062,84184..84310,
84464..84573,84672..84803,84895..85062,85184..85310,
85464..85573,85672..85803,85895..86062,86184..86310,
86464..86573,86672..86803,86895..87062,87184..87310,
87464..87573,87672..87803,87895..88062,88184..88310,
88464..88573,88672..88803,88895..89062,89184..89310,
89464..89573,89672..89803,89895..90062,90184..90310,
90464..90573,90672..90803,90895..91062,91184..91310,
91464..91573,91672..91803,91895..92062,92184..92310,
92464..92573,92672..92803,92895..93062,93184..93310,
93464..93573,93672..93803,93895..94062,94184..94310,
94464..94573,94672..94803,94895..95062,95184..95310,
95464..95573,95672..95803,95895..96062,96184..96310,
96464..96573,96672..96803,96895..97062,97184..97310,
97464..97573,97672..97803,97895..98062,98184..98310,
98464..98573,98672..98803,98895..99062,99184..99310,
99464..99573,99672..99803,99895..100062,100184..100310,
100464..100573,100672..100803,100895..101062,101184..101310,
101464..101573,101672..101803,101895..102062,102184..102310,
102464..102573,102672..102803,102895..103062,103184..103310,
103464..103573,103672..103803,103895..104062,104184..104310,
104464..104573,104672..104803,104895..105062,105184..105310,
105464..105573,105672..105803,105895..106062,106184..106310,
106464..106573,106672..106803,106895..107062,107184..107310,
107464..107573,107672..107803,107895..108062,108184..108310,
108464..108573,108672..108803,108895..109062,109184..109310,
109464..109573,109672..109803,109895..110062,110184..110310,
110464..110573,110672..110803,110895..111062,111184..111310,
111464..111573,111672..111803,111895..112062,112184..112310,
112464..112573,112672..112803,112895..113062,113184..113310,
113464..113573,113672..113803,113895..114062,114184..114310,
114464..114573,114672..114803,114895..115062,115184..115310,
115464..115573,115672..115803,115895..116062,116184..116310,
116464..116573,116672..116803,116895..117062,117184..117310,
117464..117573,117672..117803,117895..118062,118184..118310,
118464..118573,118672..118803,118895..119062,119184..119310,
119464..119573,119672..119803,119895..120062,120184..120310,
120464..120573,120672..120803,120895..121062,121184..121310,
121464..121573,121672..121803,121895..122062,122184..122310,
122464..122573,122672..122803,122895..123062,123184..123310,
123464..123573,123672..123803,123895..124062,124184..124310,
124464..124573,124672..124803,124895..125062,125184..125310,
125464..125573,125672..125803,125895..126062,126184..126310,
126464..126573,126672..126803,126895..127062,127184..127310,
127464..127573,127672..127803,127895..128062,128184..128310,
128464..128573,128672..128803,128895..129062,129184..129310,
129464..129573,129672..129803,129895..130062,130184..130310,
130464..130573,130672..130803,130895..131062,131184..131310,
131464..131573,131672..131803,131895..132062,132184..132310,
132464..132573,132672..132803,132895..133062,133184..133310,
133464..133573,133672..133803,133895..134062,134184..134310,
134464..134573,134672..134803,134895..135062,135184..135310,
135464..135573,135672..135803,135895..136062,136184..136310,
136464..136573,136672..136803,136895..137062,137184..137310,
137464..137573,137672..137803,137895..138062,138184..138310,
138464..138573,138672..138803,138895..139062,139184..139310,
139464..139573,139672..139803,139895..140062,140184..140310,
140464..140573,140672..140803,140895..141062,141184..141310,
141464..141573,141672..141803,141895..142062,142184..142310,
142464..142573,142672..142803,142895..143062,143184..143310,
143464..143573,143672..143803,143895..144062,144184..144310,
144464..144573,144672..144803,144895..145062,145184..145310,
145464..145573,145672..145803,145895..146062,146184..146310,
146464..146573,146672..146803,146895..147062,147184..147310,
147464..147573,147672..147803,147895..148062,148184..148310,
148464..148573,148672..148803,148895..149062,149184..149310,
149464..149573,149672..149803,149895..150062,150184..150310,
150464..150573,150672..150803,150895..151062,151184..151310,
151464..151573,151672..151803,151895..152062,152184..152310,
152464..152573,152672..152803,152895..153062,153184..153310,
153464..153573,153672..153803,153895..154062,154184..154310,
154464..154573,154672..154803,154895..155062,155184..155310,
155464..155573,155672..155803,155895..156062,156184..156310,
156464..156573,156672..156803,156895..157062,157184..157310,
157464..157573,157672..157803,157895..158062,158184..158310,
158464..158573,158672..158803,158895..159062,159184..159310,
159464..159573,159672..159803,159895..160062,160184..160310,
160464..160573,160672..160803,160895..161062,161184..161310,
161464..161573,161672..161803,161895..162062,162184..162310,
162464..162573,162672..162803,162895..163062,163184..163310,
163464..163573,163672..163803,163895..164062,164184..164310,
164464..164573,164672..164803,164895..165062,165184..165310,
165464..165573,165672..165803,165895..166062,166184..166310,
166464..166573,166672..166803,166895..167062,167184..167310,
167464..167573,167672..167803,167895..168062,168184..168310,
168464..168573,168672..168803,168895..169062,169184..169310,
169464..169573,169672..169803,169895..170062,170184..170310,
170464..170573,170672..170803,170895..171062,171184..171310,
171464..171573,171672..171803,171895..172062,172184..172310,
172464..172573,172672..172803,172895..173062,173184..173310,
173464..173573,173672..173803,173895..174062,174184..174310,
174464..174573,174672..174803,174895..175062,175184..175310,
175464..175573,175672..175803,175895..176062,176184..176310,
176464..176573,176672..176803,176895..177062,177184..177310,
177464..177573,177672..177803,177895..178062,178184..178310,
178464..178573,178672..178803,178895..179062,179184..179310,
179464..179573,179672..179803,179895..180062,180184..180310,
180464..180573,180672..180803,180895..181062,181184..181310,
181464..181573,181672..181803,181895..182062,182184..182310,
182464..182573,182672..182803,182895..183062,183184..183310,
183464..183573,183672..183803,183895..184062,184184..184310,
184464..184573,184672..184803,184895..185062,185184..185310,
185464..185573,185672..185803,185895..186062,186184..186310,
186464..186573,186672..186803,186895..187062,187184..187310,
187464..187573,187672..187803,187895..188062,188184..188310,
188464..188573,188672..188803,188895..189062,189184..189310,
189464..189573,189672..189803,189895..190062,190184..190310,
190464..190573,190672..190803,190895..191062,191184..191310,
191464..191573,191672..191803,191895..192062,192184..192310,
192464..192573,192672..192803,192895..193062,193184..193310,
193464..193573,193672..193803,193895..194062,194184..194310,
194464..194573,194672..194803,194895..195062,195184..195310,
195464..195573,195672..195803,195895..196062,196184..196310,
196464..196573,196672..196803,196895..197062,197184..197310,
197464..197573,197672..197803,197895..198062,198184..198310,
198464..198573,198672..198803,198895..199062,199184..199310,
199464..199573,199672..199803,199895..200062,200184..200310,
200464..200573,200672..200803,200895..201062,201184..201310,
201464..201573,201672..201803,201895..202062,202184..202310,
202464..202573,202672..202803,202895..203062,203184..203310,
203464..203573,203672..203803,203895..204062,204184..204310,
204464..204573,204672..204803,204895..205062,205184..205310,
205464..205573,205672..205803,205895..206062,206184..206310,
206464..206573,206672..206803,206895..207062,207184..207310,
207464..207573,207672..207803,207895..208062,208184..208310,
208464..208573,208672..208803,208895..209062,209184..209310,
209464..209573,209672..209803,209895..210062,210184..210310,
210464..210573,210672..210803,210895..211062,211184..211310,
211464..211573,211672..211803,211895..212062,212184..212310,
212464..212573,212672..212803,212895..213062,213184..213310,
213464..213573,213672..213803,213895..214062,214184..214310,
214464..214573,214672..214803,214895..215062,215184..215310,
215464..215573,215672..215803,215895..216062,216184..216310,
216464..216573,216672..216803,216895..217062,217184..217310,
217464..217573,217672..217803,217895..218062,218184..218310,
218464..218573,218672..218803,218895..219062,219184..219310,
219464..219573,219672..219803,219895..220062,220184..220310,
220464..220573,220672..220803,220895..221062,221184..221310,
221464..221573,221672..221803,221895..222062,222184..222310,
222464..222573,222672..222803,222895..223062,223184..223310,
223464..223573,223672..223803,223895..224062,224184..224310,
224464..224573,224672..224803,224895..225062,225184..225310,
225464..225573,225672..225803,225895..226062,226184..226310,
226464..226573,226672..226803,226895..227062,227184..227310,
227464..227573,227672..227803,227895..228062,228184..

Qy 401 LeuProArgGluGluLeuLeuArgGlyLysMetValIleCysAspArgGlyValAsnGlyArg 420
Db 63293 CTTTCCAAAGAGAGAAATCCGAGGCAAAATGGTGGATTTGGATCGGGAGTCAATGGAAGA 63352
Qy 421 SerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaSerThr 440
Db 63353 TCGGAGAAAGGAGAGCGGTTAAAGAGCTTAAAGAGCTGGAGGAGTTGCAATGATCTTACCAATACA 63412
Qy 441 GluIleAsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGly 460
Db 63413 GAGATCAACCAAGAGAGATTTCTATTGACGTTCACTCTTACCAGCTACATGATGTTGT 63472
Qy 461 TyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArg 480
Db 63473 TACACTGAGTCACTCTCTGAGGCTTATGTTAATGCCACGGTGAACCAAGGCGCG 63532
Qy 481 IleIlePheGlyThrValIleGlyAysSerArgAlaProGluValAlaGlnPheSer 500
Db 63533 ATAATTTTGGTGGTACGGTACGATGGAGGTCACGAGCACCGAGGAGTGGCTCAGTTTCA 63592
Qy 501 AlaArgGlyProSerIleAlaAsnProSerIleLeuLysProAspMetIleAlaProGly 520
Db 63593 GCTCGAGGACCGAGTTAGCAATCTCTCGATCTAAACCCGATATGATGCTCCGGA 63652
Qy 521 ValAsnIleAlaAlaTyrProGlnAsnLeuGlyProThrGlyLeuProTyrAspSer 540
Db 63653 GTCAATATCATTTGGGCTTGGCTTCAAAATCTAGGACCAACCGGACTTCTTATGATCA 63712
Qy 541 ArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGly 560
Db 63713 AGAAGAGTTAACTTCACTGTAATGTGAGGAACTTCAATGCTTGTCCCATGTGACGGA 63772
Qy 561 IleThrAlaLeuLeuArgSerAlaTyrProAsnTyrSerProAlaAlaIleLysSerAla 580
Db 63773 ATCACTGCTCTTATCCGCTGTCATCCGAACTGCTCCAGTCAATCAATCAACCGCA 63832
Qy 581 LeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaTleLysAspGlyAsnLys 600
Db 63833 TTGATGCAACACGCGATTTGTACGATCGTCAAGGAAAGCGATAAAGATGGTAAACAAA 63892
Qy 601 ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnPro 620
Db 63893 CCAGCCGGTGTGTTGCGATTTGAGCAGGCGCATGTAATCCGAAAGCGGATAACCCG 63952
Qy 621 GlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrIleuCysThrLeuGlyPhe 640
Db 63953 GGAATGGTTTACAACTTCAACCGAGTGGATTACATAACTTACCTCTGCACCTCTGGATTC 64012
Qy 641 ThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArg 660
Db 64013 ACAAGATCAGATATTTAGCAATCACTCATAGAACGTGAGCTGCAATGAATATGGCG 64072
Qy 661 LysAsnProGlyPheSerIleAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
Db 64073 AAAAACCCGGTGTAGTCTCAATTCACCGTCGATAGCGGTGATTTTCAACGTCGCAAG 64132
Qy 681 ThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSer 700
Db 64133 ACTACCGAGATGATCAACAGCGGTGCTCACTTAACGTTGGAGTCTCACTCGATATACTCA 64192
Qy 701 ValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPhe 720
Db 64193 GTGAATGTCAGGCTCCAGAGGGGATCAAGATTTATGTCATCTCAAGAGACTTGTGTTTC 64252
Qy 721 LysHisValAspGlnThrLeuSerTyrArgValThrPheValLeuLysLysLysAsnArg 740
Db 64253 AAACACGTGGATCAGCGCTAGCTATAGATGATGTTGTTATTAAGAGAAACACAGA 64312
Qy 741 GlyGlyLysValAlaSerPheAlaGlnGlyLeuThrTrpValAsnSerHisAsnLeu 760
Db 64313 GGAGGAGGTGGCTAGCTTTTGACAAGGCGACTTGACTGGTCACTCTCATATATCTG 64372
Qy 761 MetGlnArgValArgSerProIleSerValThrLeuLysThrAsn 775

Db 64373 ATGCAGCGAGTTAGAAGTCCAATCTCTGTAACCTTGAAGACTAAC 64417
RESULT 2
AC122149
LOCUS
DEFINITION
OSUNBa0083D24, complete sequence.
ACCESSION
AC122149.2 GI:29611669
KEYWORDS
HTG.
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriarthroideae; Oryzoae; Oryza.
REFERENCE
1. (Bases 1 to 150503)
Wing, R.A., Yu, Y., Soderlund, C., Kim, H.-R., Rambo, T., Currie, J.,
Collura, K., McCombie, D.W.R., de la Bastide, M., Spiegel, L.,
Preston, R., Kirchoff, K., Kuit, K., Nascimento, L., Zutavern, T.,
Balijs, V., Bell, M., Baker, J., Santos, L., Miller, B.,
Katzenberger, F., Muller, S., King, L., Yang, C., O'Shaughnessy, A.,
Palmer, L. and Bedhia, N.
Rice Genomic Sequence
Unpublished
2. (Bases 1 to 150503)
Wing, R.A., Yu, Y., Soderlund, C., Chen, M., Kim, H.-R., Rambo, T.,
Saski, C., Henry, D., Oates, R., Simmons, J., McCombie, D.W.R., de la
Bastide, M., Spiegel, L., Preston, R., Kirchoff, K., Kuit, K.,
Nascimento, L., Zutavern, T., Balijs, V., Bell, M., Baker, J.,
Santos, L., Miller, B., Katzenberger, F., Muller, S., King, L., Yang, C.,
O'Shaughnessy, A., Palmer, L. and Bedhia, N.
Direct Submission
Submitted (23-MAY-2003) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
3. (Bases 1 to 150503)
Wing, R.A., Yu, Y., Soderlund, C., Kim, H.-R., Rambo, T., Currie, J.,
Collura, K., McCombie, D.W.R., de la Bastide, M., Spiegel, L.,
Preston, R., Kirchoff, K., Kuit, K., Nascimento, L., Zutavern, T.,
Balijs, V., Bell, M., Baker, J., Santos, L., Miller, B.,
Katzenberger, F., Muller, S., King, L., Yang, C., O'Shaughnessy, A.,
Palmer, L. and Bedhia, N.
Direct Submission
Submitted (08-APR-2003) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
On Apr 8, 2003 this sequence version replaced gi:21104882.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by more than
one plasmid subclone; and the assembly was confirmed by
restriction digest. There is only PCR coverage at the following
locations: 23285-24006, 119474-119561 and 121193-121700. At bases
128212-128214 there is only single stranded transposon coverage and
at 122172-122424, 122594-122826, and 123373-123660 there is double
stranded transposon coverage. There is dinucleotide (TA) repeat at
52845-52892 of 15-24 pairs. Syntacta reads are the only coverage
at 57061-57199. There are bacterial transposons at the following
locations: 628-3459, 26704-28941, 130907-134220, 23095-24280,
118487-121172, 120153-124368, 123584-125272, 128135-130129,
127157-130772, and 130790-133540. The assembly overlaps from base
144347-150503 with OJ1172F09 (accession # AC119796). The overlap is
from 1-6156 bases on OJ1172F09. The nucleotide sequence of this BAC
clone was generated by combining Syngenta, Cold Spring Harbor
Laboratory, and Arizona Genomics Institute sequencing data.
FEATURES
Location/Qualifiers
1. 150503
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39947"
/chromosome="3"
/clone="OSUNBa0083D24"

BASE COUNT	39964 a	34575 c	33257 g	42707 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	4.45e-147	Length:	150503		
Score:	2439.00	Matches:	469		
Percent Similarity:	75.35%	Conservative:	127		
Best Local Similarity:	59.29%	Mismatches:	153		
Query Match:	60.70%	Indels:	42		
DB:	8	Gaps:	12		
US-09-806-767-2 (1-775) x AC122149 (1-150503)					
Qy	6	PhePheLeuCysIlellePheLeuLeuPheCys	-----SerSerSer	19	
Db	86354	TTCTCTGCTTTCTTCCACCTCTTCTGCTTCTGCTGTTCTTCCAGTTCTGCTGCTGCTGGA	86413		
Qy	20	SerGluIleLeuGlnLysGlnThrTyrlleValGlnLeuHisProAsnSerGluThrAla	39		
Db	86414	GCAGAGGATCTG-----CAGAGCTACATTGTGCAGCTGCATCCGATGAAGTACCTCCGT	86467		
Qy	40	LysThr-----PheAlaSerLysPheAspTrpHisLeuSerPhe	52		
Db	86468	TCCAGCGCGCAGCGCGCGAGCGGTGTTTCGGTTCGAGATGCACCTGGCATCTCTCGTTT	86527		
Qy	53	LeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSerSerArgLeuLeu	72		
Db	86528	CTTCGAGAAGTCCGTG-----GCATGGAGCGCGGAGAGCGCGTCTGCGCCTCTCTG	86581		
Qy	73	TyrSerTyrlleSerAlalleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGlu	92		
Db	86582	TATTCTTACCACACAGGTTCGTGATGCTTCGCGTGCAGCTCACGGAGGAGGAGCGCGC	86641		
Qy	93	IleLeuArgTyrlleSerProGluValAlaValArgProAspHisValLeuGlnValGln	112		
Db	86642	CGCTCGCGGAGCTCGCGCGCTCGCTCGCTCGCGCGCAGCGCGGTGAGCTGCAC	86701		
Qy	113	ThrThrTyrlleSerTyrllePheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrpSer	132		
Db	86702	ACGAGCTACTGCTACAGTTCTTGGGGCTCGAC---TTCTGTCCGACCGCGCGTGGCG	86758		
Qy	133	LysSerArgPheGlyGlnGlyThrIlelleGlyValLeuAspThrGlyValTrpProGlu	152		
Db	86759	CGGTCCGGGTACGGCGCGCACGATCATCGCGTGTCTGACACACCGCGTGTGGCGGAG	86818		
Qy	153	SerProSerPheAspThrGlyMetProSerIleProArgLysTyrlleCys	172		
Db	86819	AATCCCAAGCTTCGATGACCCCGGATGCCCGCGGTGCCCGCGGTGGCAGGCGGTGTC	86878		
Qy	173	GluGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIleGlyAlaArgPhe	192		
Db	86879	CAGGAGGTGAGCATTTCAACGCCACCACTGCAATCGGAAGCTCATCGCGCGCGGTTC	86938		
Qy	193	PheIleArgGlyHisArgVal-----AlaAsnSerProGluGluSerProAsnMetPro	210		
Db	86939	TACTCCAAAGGCGCACCGCGCAACTACCGACCAACCCCGTCCGACCGCGGTGCTGCTG	86995		
Qy	211	ArgGluTyrlleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrVal	230		
Db	86996	ATGAGGTACGTGTCGCGCGGAGCGCGCACCGGCGCGGACGACACCGCGTCCGACGCGC	87055		
Qy	231	GlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyValalalaArgGly	250		
Db	87056	GCCGGC	87115		
Qy	251	MetAlaProGlyAlaHisIleAlaValTyrlleValCysTrpPheAsnGlyCysTyrlleSer	270		
Db	87116	GTGGCCCGC	87175		
Qy	271	SerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeu	290		
Db	87176	TCCGACATCTCGCGGAGTGGACGCGCGTTCGCGGAGCGCGCGCGCGCGCGCGCGCTC	87235		
Qy	291	SerLeuGlyGlyPheProIleProLeuTyrlleAspThrIleAlaIleGlyThrPheArg	310		
Db	87236	TCCCTCGCGCGCTTCCCATCCGCTCTTTCGAGGACAGCATCGGCATCGCGCTTCGCT	87295		
Qy	311	AlaMetGluArgGlyIleSerValIleCysAlaIleGlyAsnAsnGlyProIleGluSer	330		
Db	87296	GCCACGACGATGGGTCTCCGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	87355		
Qy	331	SerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArg	350		
Db	87356	TCCGTCCGCAACGAGGCGCATGGTGATCACCTCGCGCGCGCGCGCGCGCGCGCGCGC	87415		
Qy	351	PheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrlleGlyLysLeuTyrllePro	370		
Db	87416	TTCCCGGCTATGTCGCGCTCGGGAACGCGCGGATCCTGTACGCGAGTCCATGTTCCCT	87475		
Qy	371	GlyLys---GlyIleLysAsnAlaGlyArgGluValGluValIleTyrlleValThrGlyGly	389		
Db	87476	GGGAAGTCGATTTGAAATATGGCGGAGGAGTCCGAGCTGGTGTACCGCGCGCATGGG	87535		
Qy	390	AspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLys	409		
Db	87536	ACCGCGGAGAGATGTACTGCTCAAGGGGCTCTGTCTGCCGCATGTCTGCCGGAAG	87595		
Qy	410	MetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGlu	429		
Db	87596	ATGTCGCTGTCGACCGTGGGATCACCGCGAGGCTGACAAAGCGGAGGAGTAAACAA	87655		
Qy	430	AlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIle	449		
Db	87656	GCAGCGCGCGCGCATGATTCTCGAATTCGAGATTAACACGAGGAGGAGTCCGCTC	87715		
Qy	450	AspValHisLeuLeuProAlaThrLeuIleGlyTyrlleThrGluSerValLeuLeuIle	469		
Db	87716	GATGTCCATGTCCTCGCGCTGACGCTCATCGCTACAGGGAGCGCGCTCGAGCTCAAG	87775		
Qy	470	TyrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIleGly	489		
Db	87776	TAGTCAGCTCCAGCGCGCGCGCGTGGCGAGGATAGTTCGCGCGGAGCGCGATTGGG	87835		
Qy	490	ArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnPro	509		
Db	87836	CGGCGACGCGCGCGCGGTGGCGCTGTTCTCCGCGCGCGCGCGCGCGCGCGCGCGC	87895		
Qy	510	SerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaTrpProGln	529		
Db	87896	TCGTTGCTCAAGCCGACGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	87955		
Qy	530	AsnLeuGlyProThrGlyLeuProTyrlleProTyrlleProTyrlleProTyrllePro	549		
Db	87956	AACCTGGAGCGCTCGCGCGCTGGAGGCGACCGCGCGCGCTCCGACTTCAGCGTCTCTC	88015		
Qy	550	GlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrlle	569		
Db	88016	GGGACATCAATGGGTGC	88075		
Qy	570	ProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrlle	589		
Db	88076	CCGTGCTGGAGCGCGCGCGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	88135		
Qy	590	ArgGlnGlyLysAlaIleLysAspGlyAsn---LysProAlaGlyValPheAlaIleGly	606		
Db	88136	CGGCGAGGCAAGCGCGCATGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	88195		
Qy	609	AlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrlleAsnIleGlnPro	628		
Db	88196	GCCGGGACGCTGAACCGC	88255		
Qy	629	ValAspTyrlleThrTyrlleCysThrLeuGlyPheThrArgSerAspIleLeuAlaIle	648		
Db	88256	GCCGACTAGCTACCCACCTCTGCAACCTCGGGGTACACCCACATGGAGATCTTCAAGATC	88315		
Qy	649	ThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsn	668		

```

Db 88316 ACCACGCTGGGTCAACTGACCGCGGTGGAGCGGAACGCGGTTCAGCTCAAC 88375
Qy 669 TyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArgArg 688
Db 88376 TACCCATCCATATCGGTGGGTTCAGACGAACACACACGTCGCGGTTCAGAGAGACG 88435
Qy 689 ValThrAsnValGlySerProAsnSerIleTyrSerValAsnValIysAlaProGluGly 708
Db 88436 GTACCAACGTCGCGACGCGGAACCTGACGTACACGCGCAGCTGCGCGCGCGACGCGC 88495
Qy 709 IleLysValIleValAsnProLysArgGluValPheLysHisValAspGlnThrLeuSer 728
Db 88496 GTCAGGTCAGGTGTCGCGCGCGACGCTGACGTTCTCGAGTTCGCGGAGAGAGAGC 88555
Qy 729 TyrArgValTrpPheValLeuLysLysLysAsnA-gGlyGlyValAlaSer----- 746
Db 88556 TTCGGGTG-----GCGTCGCGCGCGCGTCG 88582
Qy 747 -----PheAlaGlnGlyGlnLeuThrTrpValAsnSer---HisAsnLeu 760
Db 88583 CCGCGCGCGCACGACAAACGCGAGGGGTACCTGGTGTGGAGACGAGCGCGGAGCGGG 88642
Qy 761 MetGlnArgValArgSerProIleSerValThr 771
Db 88643 AAGCGAGGTGAGGAGCGCCCATTCGCGTGACA 88675

```

```

RESULT 3
AX506370 2265 bp DNA linear PAT 27-SEP-2002
LOCUS
DEFINITION Sequence 1065 from Patent WO0216655.
ACCESSION AX506370
VERSION AX506370.1 GI:23387607
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE
1 Harper, J.F., Krebs, J., Wang, X. and Zhu, T.
Stress-regulated genes of plants, transgenic plants containing
same, and methods of use
Patent: WO 0216655-A 1065 28-FEB-2002;
The Scripps Research Institute (US); Syngenta Participations AG
(CH)
FEATURES
source
1..2265
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 519 a 510 c 542 g 694 t
ORIGIN
Alignment Scores:
Pred. No.: 3,47e-104 Length: 2265
Score: 1742.00 Matches: 370
Percent Similarity: 63.48% Conservative: 115
Best Local Similarity: 48.43% Mismatches: 241
Query Match: 43.35% Indels: 38
DB: 6 Gaps: 15
US-09-806-767-2 (1-775) x AX506370 (1-2265)
Qy 11 IlePheLeuLeuPheCysSerSerSerGluIleLeuGlnLysGlnThrTrpIleVal 30
Db 52 CTCCTCTCTCTCTCCACACACCGCC-----AAAAAACCTCATCATC 96
Qy 31 GlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrpHisLeu 50
Db 97 CGCGTCGAAT---CACTCCGATAACCGGAATCATCTCTACTCACCAGATTGGTACACA 153
Qy 51 SerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluProSerSerArg 70

```

```

Db 154 TCTCAACTCAATTCA-----GAATCATCT 177
Qy 71 LeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGlu 90
Db 178 CTTCTCTACACTTACAAACCTCTCCATGGCTTCTCCGCTTACCTGACTCCACCGAA 237
Qy 91 AlaGlu---IleLeuArgTyrSerProGluValValAlaValArgProAspHisValLeu 109
Db 238 GCGATCTCTCTCTCTCTCTCAACTCAATCTCGATATCTTCGAAATCTCTCTCTAC 297
Qy 110 GlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGly---PheGlyAsnSer 128
Db 298 ACATTCACACTACGCGTACTCTGAGTTCTCGGTCTCAATTCGGAATTCGGTGTTCAC 357
Qy 129 GlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGly 148
Db 358 GATCTCGTTCTCTCTCT-----AAGCGGTTTATCATCGAGTTTTAGATATGGC 408
Qy 149 ValTrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrp 168
Db 409 GTTTGGCTTGAATCTAGAGCTTCGATGATCTGATGATGCTGAGATTCTTCTTAATGG 468
Qy 169 LysGlyIleCysGlnGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIle 188
Db 469 AAGGAGAAATGTAATCTGTTCCGATTTCGATTCAGTTGTAAACAAGAGCTTATC 528
Qy 189 GlyAlaArgPhePheLeuArgGlyHisArgValAlaAsnSerProGluGluSerProAsn 208
Db 529 GGAGTAGAAGCTTCTCAAAAGGATTCAAAATGGTCTCTGGTGGTGGTTCGAGTAAG 588
Qy 209 MetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSer 228
Db 589 -----CGTGAATCTGTTCTCTCTCGTGTGTTGACGGACATGGAACATCTCAACT 642
Qy 229 ThrValGlyGlySerSerValSerMetAsnValLeuGlyAsnGlyAlaGlyValAla 248
Db 643 ACCGCCCGGAGTCCGCGGTAGAACGCTAGCTTCTCGTTACGCCCGGTACGGCC 702
Qy 249 ArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys 268
Db 703 AGAGGTATGCCACTCGTCTGCTGTTGTCTATATAAGTTTGTGGAGTACTGTTGT 762
Qy 269 TyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeu 288
Db 763 TTTGGATCTGATATAGTCTGCTATGGATCGAGCTATCTGATGCTGTGTGTGCTT 822
Qy 289 SerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThr 308
Db 823 TCGTTATCTCTTGGTGGTGGTCTGCTCCGTATTATCGCATAGCATGCGATTGGAGCG 882
Qy 309 PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle 328
Db 883 TTTTCGGCTATGGAGAGAGGTGTTTGTCTGTCTGCTGTAATAGTGGTCTACT 942
Qy 329 GluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAsp 348
Db 943 AGAGCTTCTGTGCCAATGTTGCTCTGCTGTTATGACTGCTGCTGCTGCTTGTAGAT 1002
Qy 349 ArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeu 368
Db 1003 AGAGATTTTCGGGCTTTTTCGAATCTCGTAACGGGAACGACTTACCGGTGTTCCGCTG 1062
Qy 369 TyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGly 388
Db 1063 TATAGCGGTGAGAAATG-----GGACGAGCGGCTGAAATGGTTTATATAAGGG 1116
Qy 389 GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGly 408
Db 1117 AATAGTAGTTCGAGTAATCTTTGTTTACCTGGTTCGCTTGATTCAGTATTGTTGCTGGG 1176
Qy 409 LysMetValIleCysAspArgGlyValAsnGlyArgSerGlnLysGlyGluAlaValLys 428

```



```

Db 1177 AAGATTGTTGTTGTAGAGGTTGTTAATGCTAGAGTTGAGAAGAGAGCTGGTTAGA 1236
Qy 429 GluAlaGlyGlyValAlaMeIleLeuAlaAsnThrGluIleAsnGlnGluAspSer 448
Db 1237 GATGCTGCTGTTTAGGATGATAATGCGAATAGTCTGCGAGTGGAGAGAGCTTGTTG 1296
Qy 449 IleAspValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLeuLys 468
Db 1297 GCGGATAGTCAATTCCTCCCGCATCGCTAGGAGAGAGAGCTGGTGAATTTACTTAGG 1356
Qy 469 AlaTyrValAsnAlaThrValIleProLysAlaArgIleIlePheGlyGlyThrValIle 488
Db 1357 GAGTAGTAAAGTCAGATTCACCAACCCGCTCTCTGTTTAAAGAACGGTCTT 1416
Qy 489 GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsn 508
Db 1417 GACGTTAAGCCGCTCCTGCTGTTGCTGTTTAGCTCGAGAGTCTAATACCTTACT 1476
Qy 509 ProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTyrPro 528
Db 1477 CCGTAAATCTTGAAGCCTGATGTTATGCTCCTCGAGTTAATATTTGGCTGTTGCTCT 1536
Qy 529 GlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgArgValAsnPheThrValMet 548
Db 1537 GACGCTATTGTCCTACTGCTCTTCACAGGACTCTAGGAGAGCTCAGTTCAACATCATG 1596
Qy 549 SerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAla 568
Db 1597 TCAGGTAGTCAATGTCATGCCACACATCAGTGGTTAGCGGCTCTTTTGAAGCAGCT 1656
Qy 569 TyrProAsnTyrProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyr 588
Db 1657 CACCTGAGTGGAGTCCGAGTGCTATCAATCAGCTCTCAGTACAGCTTACGTTCTT 1716
Qy 589 AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla 606
Db 1717 GACACACCAACGCTCTCTCCATGATGCTCGACAGCAACAGCTTATACCCATATGCT 1776
Qy 607 IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIle 626
Db 1777 CACGCTCGGCGCATGATGATCCCAAGAGCTCTCCACCAAGCTCTTGCTAGCAGTCT 1836
Qy 627 GlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeu 646
Db 1837 TCACCGGAGGAATACATCAGGTTTTTGTGCTCTCTAGACTACACAGTCGATCAGTTGTT 1896
Qy 647 AlaIleThrHisLys-----AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPhe 665
Db 1897 GCGATTGTGAAGCAGCACCTAGCGTTAACTGCTCGAAGAAGTTC---TCAGATCCTGCT 1950
Qy 666 SerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIle 685
Db 1951 CAGCTCAACTACCAAGTTCTCGGTTTGTGTT---GGGGGTAAAGAGAGTTGTCGGTAC 2007
Qy 686 ThrArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAla 705
Db 2008 ACTCGGAGAGTACAAATGTTGTCGACAGCTCGGTTTACAAAGTACGCGTTAATGGA 2067
Qy 706 ProGluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGln 725
Db 2068 GCTCTAGTGTGCGAATCTCTGTTAAACCATCGAACTTTCGTTTAAAGCGTGGGAGAG 2127
Qy 726 ThrLeuSerTyrArgValTyrPheValLeuLysLysLysAsnArgGlyGlyLysValAla 745
Db 2128 AAGAGAGAGGTACACAGTCACGTCACGTTTGTGTAGCAAGAAGAGAGTGTAGTACGAACAGGCT 2187
Qy 746 SerPheAlaGlnGlyGlnLeuThrTyrValAsnSerHisAsnLeuMetGlnArgValArg 765
Db 2188 GAGTTT-----GGTTCCATCCTTGAGCAATCCGAGCAC-----GAAGTGAGA 2232
Qy 766 SerProIleSer 769
Db 2233 AGTCCCGTTGCA 2244

```

RESULT 4

AY142613

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

CDS

AY142613 2296 bp mRNA linear PLN 23-SEP-2002
Arabidopsis thaliana putative subtilisin serine protease
(At2g05920) mRNA, complete cds.

AY142613.1 GI:23296837
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (bases 1 to 2296)
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W.,
Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K.,
Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R.,
Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J.,
Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M.,
Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.

Arabidopsis Open Reading Frame (ORF) Clones
Unpublished
2 (bases 1 to 2296)

Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W.,
Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K.,
Wong, C., Wu, H.C., Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R.,
Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J.,
Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M.,
Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.

Submitted (16-AUG-2002) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA

The RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-Length cDNA': Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and
sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada, K.,
Chan, M.M., Chang, C.H., Dale, J.M., Deng, J.M., Hsuan, V.W., Lee, J.M.,
Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C.,
Wu, H.C., Yu, G., Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J.,
Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Tripp, M.G., Wu, T.,
Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally
to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP
/PGEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to GenBank.

Location/Qualifiers

1..2296
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="U09230"
/note="This clone is in pUNI 51.
ecotype: Columbia"
1..2296
/gene="At2g05920"
1..2265
/gene="At2g05920"
/codon_start=1
/evidence=experimental
/product="putative subtilisin serine protease"
/protein_id="AA113182.1"
/db_xref="GI:23296838"

/translation="WASSSSSSSSIIITITFLILLHTTAKKTYIIRVHSDKPEFL
THDWYSQLNSESLLYITTSFPHGSAYLSDSTEADSLSSNSLIDIFEDPLYTLH
TRTFPELGLNSEFEVHDLSSNGVILGLDTGVMPESRSFDTPMPPIPSKWKGE
ESGDFPSKLCNKKLIGARFSKGFQWASGGFSKRESVSRDVGHGHTSTTAAG
SAVNASGLYAACTGATRAVATYKVCWSTGCFGSDILAAMDRAILLDGVVLSL
SLGGSPAYEDTIAICAFSAMERGVEFVSCAGNSGPTASVANVAPWMTVAGATLD
RDPFAPNLNGKRLTGVSILYSGVGMGTKELELYVYKNGSSSNLCLPSLSSIVRG
KIVCDGVNARVEKAVRDAGGLGMANTAAAGEELVADSHLLPALTAVGKKTGDL
LREYVSDSPTALLVFKGTVDLVKPSVVAASFSSRGPNTVPEILKPDVPGVNIIL
AGWSDAIGPTGLDKDSRTOFNIMSGTSMSCPHISGLAGLLKAAHPWSPSAIKSALM
TAVVLDNAPLHDAANDSLNPSYAHSGHVDPOKALSGLVVDI,STREYIRFLCSL
DYTVDHVAIVKRVNCSKFKSPDGLNLYPESVLFGKRVVYTRVNTVGAASSV
YKVTNAPSVGISVKPSKLSFKSVGEKKRYITVTVSKGVSMINKAEPGSIWSPQ
HEVRSVPVAFSNRF"
2266..2296
/Gene="At2g05920"
3' UTR
BASE COUNT 532 a 515 c 546 g 703 t
ORIGIN

Alignment Scores:
Pred. No.: 3,54e-104 Length: 2296
Score: 1742.00 Matches: 370
Percent Similarity: 63.48% Conservative: 115
Best Local Similarity: 48.43% Mismatches: 241
Query Match: 43.35% Indels: 38
DB: 8 Gaps: 15

US-09-806-767-2 (1-775) x AV145613 (1-2296)

Qy 11 IlePheLeuLeuPheCysSerSerSerGluIleLeuGlnLysGlnThrTyrlleVal 30
Db 52 CTCTCTCTCTCTCCACCAACGCC-----AAAAAACCTACATCATC 96
Qy 31 GlnLeuHisProAenSerGluThrAlaLysThrPheAlaSerLysPheAspTrpHisLeu 50
Db 97 CGGTCGAAT---CACTCCGATAAACCGGAATCATCTCTCACTCACCGAGATTGGTACACA 153
Qy 51 SerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSerSerArg 70
Db 154 TCTCAACTCAATTCA-----GNAATCATCT 177
Qy 71 LeuLeuTyrserTyrglySerAlaIleGluGluPheAlaAlaGlnLeuThrGluSerGlu 90
Db 178 CTCTCTACACTTACACAACTCTCTCCATGGCTTCTCCGCTTACCTCGACTCCACCGAA 237
Qy 91 AlaGlu---IleLeuArgTyrsProGluValValAlaValArgProAspHisValLeu 109
Db 238 GCGGATTCCTCTCTCTCTCAAACTCAATCTCGATATCTTCGAAGATCTCTCTAC 297
Qy 110 GlnValGlnThrThrTyrsTyrsPheLeuGlyLeuAspGly---PheGlyAsnSer 128
Db 298 ACATCTCACACTACGCGTACTCTCGAGTTCTCGGTCTCAATTCGGAATTCGGTTCAC 357
Qy 129 GlyValTrpSerLysSerArgPheGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 148
Db 358 GATCTCGGTTCTTCTCT-----AACGGCGTTATCATCGAGTTTATGATCTGCG 408
Qy 149 ValTrpProGluSerProSerPheAspAspThrGlyWetProSerIleProArgLysTrp 168
Db 409 GTTTGGCTGAATCTAGAGCTTCGATGATACATGATATGCTGAGATTCCTCTTAATGG 468
Qy 169 LysGlyIleCysGlnGlnGluGluSerPheSerSerSerCysAsnArgLysLeuIle 188
Db 469 AAGAGAGATGTAATCTGTTCCGATTTCCGATTTCCGATTTGTTAACGAAGCTTATC 528
Qy 189 GlyAlaArgPhePheIleArgGlyHisArgValAlaAlaAsnSerProGluGluSerProAsn 208
Db 529 GGAGCTAGAAGCTTCTCCAAAGGATTTCAATGCTTCTGGTGGTGGTGGTGGTGGTGGT 588
Qy 209 MetProArgGluTyrlleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSer 228
Db 589 -----CGTGAATCTGTTTCTCTCGTGTGTTGTCGAGCATGCGAATCATCTTCAACT 642

Qy 229 ThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAla 248
Db 643 ACCGCCGGGGATCCCGCTAGAACCGCTAGCTTCTCGGTACCGCCCGCGTACGCC 702
Qy 249 ArgGlyMetAlaProGlyAlaHisIleAlaValTyrlsValCysTrpPheAsnGlyCys 268
Db 703 AGAGTATGGCCACTCGTGTCTGCTTACTTATAAAGTTTGTGGAGTACTGGTGT 762
Qy 269 TyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLysValAspValLeu 288
Db 763 TTTGGATCTGATATAGCTGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 822
Qy 289 SerLeuSerLeuGlyGlyPheProIleProLeuTyrsAspThrIleAlaIleGlyThr 308
Db 823 TCGTTATCTCTGTTGGTGGTCTCTGCTCGGTATATCGGATACGATTCGATTCGAGCG 882
Qy 309 PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle 328
Db 883 TTTTCGGCTATGGAGAGAGGTGTTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 942
Qy 329 GluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyValAlaGlyThrLeuAsp 348
Db 943 AGAGCTTCTGTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1002
Qy 349 ArgArgPheProAlaValArgLeuAlaAsnGlyLysLeuLeuTyrglyGluSerLeu 368
Db 1003 AGAGATTTTCGGCTTTTTCGAACTCGTAAACGGGAAACGACTTACCGGTGTTTCGCTG 1062
Qy 369 TyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrlleValThrGly 388
Db 1063 TATACGGCTGAGGAATG-----GGGACGAGGCGCTTGAATTTGTTTATAATAAGGG 1116
Qy 389 GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGly 408
Db 1117 AATAGTAGTTCGAGTAATCTTTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1176
Qy 409 LysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLys 428
Db 1177 AAGATTGTTGTTGATAGAGGTGTTAATGCTAGAGTTGAGAAAGAGGCTGCTGTAGA 1236
Qy 429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSer 448
Db 1237 GATGCTGCTGTTTGGGATGATAATGCGAATCTGCTGCGAGTGGAGAGGAGCTGTGTG 1296
Qy 449 IleAspValHisLeuLeuProAlaThrLeuIleGlyTyrlleGluSerValLeuLys 468
Db 1297 CCGGATAGTCTATTGCTTCCGCGATCGCTGAGGAGAGAGTGTGTGATTTACTTAGG 1356
Qy 469 AlaTyrlleValAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIle 488
Db 1357 GAGTATGTTAAGTCAGATTCTAAACCAACCGCTCTTCTGTTTAAAGGAACGGTCTTT 1416
Qy 489 GlyArgSerArgAlaProGluValAlaAlaGlnPheSerAlaArgGlyProSerLeuAlaAsn 508
Db 1417 GACGTTAAGCGCTCTCTGTTGTTGCTGCTTTAGCTCGAGAGGCTCTAATATCTGTTACT 1476
Qy 509 ProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpPro 528
Db 1477 CCTGAATCTTTGAAGCCTGATGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1536
Qy 529 GlnAsnLeuGlyProThrGlyLeuProTyrsAspSerArgValAsnPheThrValMet 548
Db 1537 GACGCTATTGTTCTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1596
Qy 549 SerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAla 568
Db 1597 TCAGTACGTCATGTCATGCCACACATCAGTGGTTAGCGGGTCTTTTGAAGAGAGCT 1656
Qy 569 TyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrl 588
Db 1657 CACCTGAGTGGAGTCCGAGTGTCTATCAATCAGCTCTCATGCTACAGCTTACGTTCTT 1716
Qy 589 AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla 606

```

Db 1717 GACACACCAAGCTCTCTCCATGATGCTGCGACACAGCGCTATCTACCAATAGCT 1776
Qy 607 ILeGlyAlaGlyHisValAsnProGlnGlyLeuAlaLeAsnProGlyLeuValTyAsnIle 626
Db 1777 CACGGCTCGGCCATGATGATCCCAAAAGGCTCTCTACCAAGGCTCTGCTAGGACATC 1836
Qy 627 GlnProValAspTyIleThrTyLeuGlyThrLeuGlyPheThrArgSerAspIleLeu 646
Db 1837 TCACACGAGGAATACATCAGGTTTTTGTGCTCTCTAGACTACACAGCTCGCATCTGT 1896
Qy 647 AlaIleThrHisLys---AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPhe 665
Db 1997 GCGATTCTGAAGCGACCTAGCGTTAACTGCTCGAAGAAGTTC---TCAGATCTCGT 1950
Qy 666 SerLeuAsnTyProSerIleAlaValIlePheLysArgGlyLysThrGluMetIle 685
Db 1951 CAGTCACTACCAAGTTCTCGGTTTCTTT---GGGGTAAAGAGTTGTCGGGTAC 2007
Qy 686 ThrArgValThrAsnValGlySerProAsnSerIleTySerValAsnValLysAla 705
Db 2008 ACTCGGGAAGTAACAATGTTGTGTCAGCAAGCTCGGTTTACAAAGTGACGGTTAATGGA 2067
Qy 706 ProGluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGln 725
Db 2068 GTCCTAGTGTGGAATCTCTGTAAACCATCGAAACTTCGTTTAAAGCGTGGAGAG 2127
Qy 726 ThrLeuSerTyArgValTyrPheValLeuLysLysAsnArgGlyGlyLysValAla 745
Db 2128 AAGNAGAGGTACACATGACGTTTGTAGCAAGAAAGGAGTGTAGTACGACAAAGGCT 2187
Qy 746 SerPheAlaGlnGlyLeuThrTyrValAsnSerHisAsnLeuMetGlnArgValArg 765
Db 2188 GAGTTT-----GGTTCATCACTTGAGCAATCCGAGCAC-----GAAGTGAGA 2232
Qy 766 SerProIleSer 769
Db 2233 AGTCCCGTGTGCA 2244

RESULT 5
AY035090
LOCUS AY035090 2541 bp mRNA linear PLN 18-SEP-2002
DEFINITION Arabidopsis thaliana putative subtilisin serine protease
            (At2g05920) mRNA, complete cds.
ACCESSION AY035090.1 GI:14334833
VERSION AY035090.1
KEYWORDS FULL CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 2541)
AUTHORS Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Shinn,P., Southwick,A.,
            Satou,M., Seki,M., Shinn,P., Shinozaki,K., Sakurai,T.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
TITLE Arabidopsis Full Length cDNA Clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2541)
AUTHORS Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G.,
            Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
            Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
            Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and
            Theologis,A.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2001) Plant Gene Expression Center, 800 Buchanan

```

Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
Arabidopsis Full-Length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,
Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D.,
Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L.,
Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Koesema,E.,
Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J.,
Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGECC) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGECC)
contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to GenBank.

FEATURES

source 1. 2541

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="2"

/clone="RAFL09-15-A05 (R09230)"

/note="This clone is in a modified pBluescript vector
(FLC-1) as a BamHI/XhoI insert."

ecotype: Columbia"

gene 1. 2541

/gene="At2g05920"

5'UTR 1. 22

/gene="At2g05920"

misc_difference 2

/gene="At2g05920"

/note="compared to genomic sequence"

/replace="C"

CDS 23. 2287

/gene="At2g05920"

/codon_start=1

/evidence="experimental"

/product="putative subtilisin serine protease"

/protein_id="AAK59595.1"

/db_xref="GI:14334834"

/translation="MASSSSSSSITITFTFLHLLHTAKTYIIRVNSDKPESPL
THDWTSQLNSSESLYTYTFHGFSAVLDSTEADSLSSNSILDIPEDPLTYLH
TTRPEFLGNSFPGVHDLGSSNGVIGVLDTGVPESRSPDTPMPELPSKWKGE
ESGSDPDKLCKNKLIGARSPKGFQMGSGGFSKRESVSPDVGHGHTTTTAAG
SAVNASPLGAACTAGMATRATRVATYKVCWSTGCGSDILAAMDRAILDGVVLSL
SLGGSAPYTRDTIAGFAMERGVFVSCAGNSGPTRASVANVAPWVMTVAGLTD
RDPPAFANLGNKRLTGVSLSYGVGMGTKELELVYKNGSSSNLCLPGSLDSSI
KIVVDGVARVEKGVADAGGLGMINANTAAAGELVADSHLLPATAVGGKTGL
LREVDGSKPTALLVFKGLDMSKPTNFVMSKPSVAFSSGPNVTTPILKPDVPGVNL
AGWSDAIGTGLDKDPTNFVMSKPSVAFSSGPNVTTPILKPDVPGVNL
TTAYLNTNAPLHDAADNLSNPYAGSHVDQKALSPGLYDSTERYIFLCSL
DYTVDHVALVKKPSVNCSSKSKFSKLSKFSVGEKRYTFTVSKGVSMTNKAEGSITWSNPQ
HEVSPVPAFWSNR"

3'UTR 2288. 2541

/gene="At2g05920"

misc_difference 2524. 2525

/gene="At2g05920"

/note="not present in genomic sequence"

/note="not present in genomic sequence"

BASE COUNT 617 a 549 c 600 g

ORIGIN

Alignment Scores:

Pred. No.: 4.07e-104 Length: 2541

Score: 1742.00 Matches: 370

Percent Similarity: 63.48% Conservative: 115

Best Local Similarity: 48.43% Mismatches: 241


```

organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/variety="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosomes="4"
/clone="OSJNBa0019K04"
/clone_lib="CUGI-OSJNBa"
BASE COUNT   41665 a 35105 c 35893 g 41997 t    203 others
ORIGIN

Alignment Scores:
Pred. No.:      2.3e-99          Length:      154863
Score:           1706.50         Matches:       353
Percent Similarity: 63.23%       Conservative:  137
Best Local Similarity: 45.55%     Mismatches:   258
Query Match:      42.47%         Indels:        27
DB:                2              Gaps:            14

US-09-806-767-2 (1-775) x OSJN00071 (1-154863)

Qy      8 LeuCysTleIlePheLeuLeuPheCysSerSerSerSerGluIleu---GlnLysGln 26
Db      66607 CTGGCTGTGGTACTGTCTTTCCCAAGCAAGCGCTCTCTGCTTGCGGCCAACCTCCCAG 66548
Qy     27 ThrTyriLeValGlnLeuHisProAsnSerGluThrAlaLyThrsPheAlaSerLysPhe 46
Db     66547 ACTTACATCTGCAGATG--GCGGCGTCTGAGATGCCGAGCTCGTCGATTTTACCAT 66491
Qy     47 AspTrpHisLeuSerPheLeuGlnGluAlaValLeuGly---ValGluGluGluGluGlu 65
Db     66490 GAGTGTGTACCTTCCACGCTGAAGTCGGTGAGCTCTTCGCAGCTCGAAGATGAAGAGAT 66431
Qy     66 GluProSerSerArgLeuLeuTySerTyGlySerAlaIleGluGlyPheAlaLaGln 85
Db     66430 GATGCTTCCACGAGGATTATCATCACTACGAGAGCGGATTCATCGGTTCCGGCTCAG 66371
Qy     86 LeuThrGluSerGluAlaGluIleLeuArgTySerProGluValValAlaValArgPro 105
Db     66370 CTCGACGAGGAGAAGCCGAGCTGATGCGCGAGCGGCGGCTGCTCGCGTGTATCCCG 66311
Qy    106 AspHisValLeuGlnValGlnThrThrTySerTyLysPheLeuGlyLeuAspGlyPhe 125
Db    66310 GAGACGGTGTGAGTGCACACACACGAGAGCCGGACTTCTCGCATTTGGCCCGGAG 66251
Qy    126 GlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeu 145
Db    66250 GTCAACAACAGGATATGTCGACACGCTCGCCGACACGACGTCGTCTCGCGCTGTCTC 66191
Qy    146 AspThrGlyValTrpProGluSerProSerPheAspThrGlyWetProSerIlePro 165
Db    66190 GACACCGGATATGCGCGGAGAGCCCCAGCTTCACGACAGGGGCTCGGCCCGTGC CG 66131
Qy    166 ArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArg 185
Db    66130 GCCAAGTGAAGGCTCTGCGCAGACCGCGCGGCTTCCACACGCGGCACTGCAACCGC 66071
Qy    186 LysLeuIleClyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGlu 205
Db    66070 AAGATCGTGGGCGCGCATCTTCTACACGCGTACGAGGCGTCTGCTGGGCGCC----- 66017
Qy    206 SerProAsnMetProArgGluTyriLeSerAlaArgAspSerThrGlyHiGlyThrHis 225
Db    66016 ---ATCAACGAGACGACCGAGCTCAAGTCCCGCGCACACGAGGACGCGGACCGGACCGCAC 65960
Qy    226 ThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAla 245
Db    65959 ACCGCGCCACCGCGCGGCTCGCCCGTGAGGAGCGCCAACTGTACGGGTACGCCGCGC 65900
Qy    246 GlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyriLysValCtyrPhe 265
Db    65899 GGCGTCCCGCTGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 65840

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/variety="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosomes="4"
/clone="OSJNBa0019K04"
/clone_lib="CUGI-OSJNBa"
BASE COUNT   41665 a 35105 c 35893 g 41997 t    203 others
ORIGIN

Alignment Scores:
Pred. No.:      2.3e-99          Length:      154863
Score:           1706.50         Matches:       353
Percent Similarity: 63.23%       Conservative:  137
Best Local Similarity: 45.55%     Mismatches:   258
Query Match:      42.47%         Indels:        27
DB:                2              Gaps:            14

US-09-806-767-2 (1-775) x OSJN00071 (1-154863)

Qy      8 LeuCysTleIlePheLeuLeuPheCysSerSerSerSerGluIleu---GlnLysGln 26
Db      66607 CTGGCTGTGGTACTGTCTTTCCCAAGCAAGCGCTCTCTGCTTGCGGCCAACCTCCCAG 66548
Qy     27 ThrTyriLeValGlnLeuHisProAsnSerGluThrAlaLyThrsPheAlaSerLysPhe 46
Db     66547 ACTTACATCTGCAGATG--GCGGCGTCTGAGATGCCGAGCTCGTCGATTTTACCAT 66491
Qy     47 AspTrpHisLeuSerPheLeuGlnGluAlaValLeuGly---ValGluGluGluGluGlu 65
Db     66490 GAGTGTGTACCTTCCACGCTGAAGTCGGTGAGCTCTTCGCAGCTCGAAGATGAAGAGAT 66431
Qy     66 GluProSerSerArgLeuLeuTySerTyGlySerAlaIleGluGlyPheAlaLaGln 85
Db     66430 GATGCTTCCACGAGGATTATCATCACTACGAGAGCGGATTCATCGGTTCCGGCTCAG 66371
Qy     86 LeuThrGluSerGluAlaGluIleLeuArgTySerProGluValValAlaValArgPro 105
Db     66370 CTCGACGAGGAGAAGCCGAGCTGATGCGCGAGCGGCGGCTGCTCGCGTGTATCCCG 66311
Qy    106 AspHisValLeuGlnValGlnThrThrTySerTyLysPheLeuGlyLeuAspGlyPhe 125
Db    66310 GAGACGGTGTGAGTGCACACACACGAGAGCCGGACTTCTCGCATTTGGCCCGGAG 66251
Qy    126 GlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeu 145
Db    66250 GTCAACAACAGGATATGTCGACACGCTCGCCGACACGACGTCGTCTCGCGCTGTCTC 66191
Qy    146 AspThrGlyValTrpProGluSerProSerPheAspThrGlyWetProSerIlePro 165
Db    66190 GACACCGGATATGCGCGGAGAGCCCCAGCTTCACGACAGGGGCTCGGCCCGTGC CG 66131
Qy    166 ArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArg 185
Db    66130 GCCAAGTGAAGGCTCTGCGCAGACCGCGCGGCTTCCACACGCGGCACTGCAACCGC 66071
Qy    186 LysLeuIleClyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGlu 205
Db    66070 AAGATCGTGGGCGCGCATCTTCTACACGCGTACGAGGCGTCTGCTGGGCGCC----- 66017
Qy    206 SerProAsnMetProArgGluTyriLeSerAlaArgAspSerThrGlyHiGlyThrHis 225
Db    66016 ---ATCAACGAGACGACCGAGCTCAAGTCCCGCGCACACGAGGACGCGGACCGGACCGCAC 65960
Qy    226 ThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAla 245
Db    65959 ACCGCGCCACCGCGCGGCTCGCCCGTGAGGAGCGCCAACTGTACGGGTACGCCGCGC 65900
Qy    246 GlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyriLysValCtyrPhe 265
Db    65899 GGCGTCCCGCTGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 65840

RESULTS
LOCUS       OSJN00071/c                      154863 bp    DNA             linear      HTG 17-JUL-2002
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 4 clone
VERSION     AL606640.2 GI:21740962
KEYWORDS    HTG; HGSC PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM   Oryza sativa (japonica cultivar-group)
REFERENCE  1 Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.Q., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L., Lu X.T., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y.X., Li C., Li T., Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y., Hu Q.P., Zhang X.B., Zhang W., Wang L.J., Ding C.W., Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H. and Hong G.F.
Direct Submission
Submitted (27-JUN-2002) Han Bin, National Center for Gene Research, Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233, CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests: bhan@ncgr.ac.cn
REMARK     Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC clone: OSJNBa0019K04.
COMMENT    On Jul 12, 2002 this sequence version replaced gi:15594099.
Web site: http://www.ncgr.ac.cn
----- Summary Statistics -----
Assembly program: phrap

NOTE: This is a PHASE2 sequence. Gaps are shown by nnnnn. Genes were identified by a combination of several methods: Gene prediction programs including Fgenesh (http://www.softberry.com/), genscan (http://ccr-081.mit.edu/GENSCAN.html), GeneMarkES (http://genemark.biology.gatech.edu/GeneMark/), tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/), searches of the complete sequence against NCBI non-redundant protein database (nr) (ftp://ncbi.nlm.nih.gov/blast/db) and the EST database at NCGR.
* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 6953: contig of 6953 bp in length
* 6954 70153: gap of 200 bp
* 70154 154863: contig of 84710 bp in length.
Location/Qualifiers
1..154863
```

QY 266 AsnGlyCysTyrSerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLysVal 285
 Db 65839 GGAGGGTGTCTAGCTCCGACATCTGGCGCGCTGATCGCGCGTGTGGAGCGGCTC 65780
 QY 286 AspValLeuSerLeuSerIleGlyPheProIleProLeuTyrAspAspThrIleAla 305
 Db 65779 GAGTGTCTCTCCATCTCTCCCGGGGTGGCGCTCCCGGTACTACCTCGACAGCTGTCC 65720
 QY 306 IleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsn 325
 Db 65719 ATAGCTGTCTGGGGCCATGAGATGGCGTGTCTGCTGCTGCTGCGCGCGCAACGCC 65660
 QY 326 GlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGly 345
 Db 65659 GGGCTACCGGATAGCCCTCACCACTGTGCGCGTGGATACCAACGAGTGGGCGGAGC 65600
 QY 346 ThrLeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuTyrGly 365
 Db 65599 ACCATGACCGTGACTTCCCGGCGACGCTCGCAACGCGCGCAACATCAACCGGC 65540
 QY 366 GluSerLeuTyrProGlyLysGlyIleLysAsnAlaGly-----ArgGluValGluVal 383
 Db 65539 GTCTCGGCTTAC-----AAAGCGCTCGGAATCTGTGCGCGCGAGGAGCATCCGGTG 65486
 QY 384 IleTyrValThrGlyGlyAspLysGlySerGlu-----PheCysLeuArgGlySer 400
 Db 65485 GTCTACTTGGGGCAATTCAGCATCGCGGACCCCGCTCTCTGTGCTCGCGGGGAGC 65426
 QY 401 LeuProArgGluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArg 420
 Db 65425 CTGACGCGCACGACGCTCCCGGAGATAGTGTGACCGCGGCATCACTGCTCGG 65366
 QY 421 SerGluLysGlyGluAlaValLysGlyAlaGlyValAlaMetIleLeuAlaAsnThr 440
 Db 65365 GTGCAAGAGGTGATGGTTGCAAGAGCGCGGATCGCATGATATCTGCCAACACC 65306
 QY 441 GluIleAsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGly 460
 Db 65305 CGCGCAACCGGAGGAGCTGTGCGCGACGACCTGTGCGCGCGCTGCGAGTCCGC 65246
 QY 461 TyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArg 480
 Db 65245 GAGCGGAGGAGCATTCGCGCAAGAGCTACAGCAAGAGCGCCCAAAACCGACCGGAGC 65186
 QY 481 IleIlePheGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSer 500
 Db 65185 CTCAGCTTCGCGGAGAACCACTCGCATCCGCGCTGCGCGTGTGCTGCGCGTCTCG 65126
 QY 501 AlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGly 520
 Db 65125 TCCGCGGAGCAACATCTGATCTGGAGATCTCAAGCGGAGCTGTCGCGCGCGC 65066
 QY 521 ValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSer 540
 Db 65065 GTGAACATCTGCGGATGAGCGCGGAGCGCGGCGCGCTGCGAGTGTCCAGCGACGC 65006
 QY 541 ArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGly 560
 Db 65005 CGTCCGCTGGGTTCACTCTGTCGGGAGCTCCATGCTGCGCGCGCGACGTCGCGCGC 64946
 QY 561 IleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAla 580
 Db 64945 GTGCGCGCTGTATCAAGCGGAGCGCCCGGATGAGGAGCGCGGAGTCAAGTCCGCG 64886
 QY 581 LeuMetThrThrAlaAspLeuTyrAspArgGlnGlyAlaIleLysAspGly----- 598
 Db 64885 CTGATGACCGCGGTACTGTCAGCAACACATGACGCGCGCGGATGAGGAGCGCGGAC 64826
 QY 599 AsnLysProAlaGlyValPheAlaIleGlyValAlaGlyHisValAsnProGlnLysAlaIle 618
 Db 64825 GGCAAGCGCTCGACCGCGTTCGAGCAGCGGAGCGGCGACATACACCGCGTGGAGCTCTC 64766
 QY 619 AsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeu 638

Db 64765 ACCCTGCGCTGTCTAGCATTTGGCCAGCCGACTACCTGGAGTTCTCTGTCACGCAG 64706
 QY 639 GlyPheThrArgSerAspIleLeuAlaIleThr-----HisLysAsnValSerCysAsnGly 657
 Db 64705 CACATGACCGCGATGCGAGCTCAGGACCTTCAACAGAACTCCCAACATGATGACGCGAC 64646
 QY 658 IleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLys 677
 Db 64645 ACCTTC-----AGTTCGCGCGAGCCTCACTATCCGCGCATCTCAGTCGTCTTCGCC 64592
 QY 678 ArgGlyLysThrThrGluMetIleThrArgArg-----ValThrAsnValGlySerProAsn 696
 Db 64591 GACCAAGCATCCAAAGCGCTGACGTCGCGCGACTGTCAAGCAAGCTGTGCTCTCCATCT 64532
 QY 697 SerIleTyrSerValAsnValLysAlaProGluLysValIleLysValIleValAsnProLys 716
 Db 64531 TCGACGATACCATGTCACAGGTCACCAAGTCAAGGCGCGCGATGTCATCGTCAACCGAAC 64472
 QY 717 ArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTyrPheValLeuLys 736
 Db 64471 ACCTGTCATCTGTAAGCAACACCAAGTTAGCTTAAAGTTACAGTACAGTACACCAACAG 64412
 QY 737 LysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGlyLysLeuThrTrpValAsn 756
 Db 64411 GCGCGCCAG-----AAGGCACCGAGTTC-----GGAGCCTGTCTCTGAGCGGAC 64367
 QY 757 SerHisAsnLeuMetGlnArgValArgSerProIleSerValThr 771
 Db 64366 GGCCTCCACATT-----GTCAGAGCGCTGTGCTGCTCTAAACA 64331

RESULT 7
 AB022220/c AB022220 83906 bp DNA linear PLN 27-DEC-2000
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MLN21.
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MLN21.
 ACCESSION AB022220 BA000014
 VERSION AB022220.1 GI:4159709
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 AUTHORS Sato, S., Nakamura, Y., Kaneko, T., Kato, T., Asamizu, E. and Tabata, S.
 TITLE Structural analysis of Arabidopsis thaliana chromosome 3. I.
 JOURNAL Sequence features of the regions of 4,504,864 bp covered by sixty
 MEDLINE p1 and TAC clones
 PUBMED DNA Res. 7 (2), 131-135 (2000)
 REFERENCE 20277480
 10819329
 2 (bases 1 to 83906)
 AUTHORS Sato, S., Nakamura, Y., Kaneko, T., Kato, T., Asamizu, E. and Tabata, S.
 TITLE Direct Submission
 JOURNAL Submitted (08-JAN-1999) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research, 1332-3, Yana,
 Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
 Tel: 81-438-52-3935, Fax: 81-438-52-3934)
 COMMENT Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/sgd_graph.cgi?c=MLN21
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/Grail-1.3/),
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://gremmlin.zool.iastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE

(Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/trnascan-se/>). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MLE3 and the 3' clone is MOA2.

FEATURES

source

```

1..83906
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MLN21"
/clone_lib="Mitsui P1"
complement(1..91)
/product="GDSL-motif lipase/hydrolase-like protein"
/notes="CDS is reported in Acc# AP000416
gene id: MLE3.2"
/number=1
/evidence=not_experimental
complement(join(1714..2552,2810..3098))
/notes="gene id: MLN21.1"
/codon_start=1
/evidence=not_experimental
/product="transcription factor EREBP-like protein"
/protein_id="BAB01029.1"
/db_xref="GI:9279571"
/translations="MCGGAIISDFIPPRSLRVTFNEFIWPDLLKNVKASKRSNKRSD
FPDLDDFADFGQDDSAFDCEDDDYFVNKVPFTATTKPFASEAFVTGSGAYA
KXIVSAEQHEKSKRRKQKIGIRQRFQWGAARIPRQKSGNLTGFTDABEAA
RAYDAARINRFGKAVNPFKEKPSVQKPSAKTNLQKSAKPNKSVTLVQOPT
HLQQCCNNSFDNSFGDMFMEKPMQYNNQFLTNSFDAGNGNGYQYFSSDQGNSE
DCFEWGHGPTPISSMLVNNNEASFVEETNAAKLKPNSSDDLLMAYLDNALW
DTPLEWMLGADAGATQEEENPVELMSLDEINFMLEGDF"
complement(5735..8062)
/notes="gene id: MLN21.2"
/codon_start=1
/evidence=not_experimental
/product="subtilisin proteinase-like protein"
/protein_id="BAB01030.1"
/db_xref="GI:9279572"
/translations="MAFFYFFLLTSSPSSASSNSLTVIVHVDHEAKPSIFPTH
HTYTSLSASLTSPSIIHTYDVFHFSARLTSQDASQLLDHPHVSIVPEQVRLH
FHTSPFEGLRSTQAGLLESDFDGLVIGDVTGWPERPSPDRLGLGPVPIKWK
GQICASQDPESACNKLKVGARFCGGYEATNGKNETTEFRSPRSDGHHGHTASIS
AGRYVPASTLGVAHGVAAGMAPKARLAKYKVCWNSGCCYDSILAPATVADGVDVI
SLSVGVVVPYIDAIFAGFAIDRGIFFVSASAGNGPCALTITNAPWMTTVGAGT
IDRFANVKGKNGKMSIGSVYVGGPLDPRMYPLVYGGSLGDEGSSSLCEGLS
DPLNVKGKIVLDRGINSRATKEIVRKNGLGMIITANGFDGGLVADCHVLPATSV
GASGGBEIRRYIESKSRSSKHTATIVFKGRLGIRPAPVAVGSFARGPNPTEPI
LKPDVITAPGNILAAWPDRIGPSVTSNRTTEENTLSTSNACHVSLAALLAAH
PDWSPAIRSAITATYVDNNGEPMWDESTNGTSSVMDYGSGHVHTKAMPDGLVID
ITSYDINFLCNSYTRTNVITIRQACDCCARRAGHVGYNLNPFSFVYFOQGESK
MSTHTRITVNGDSVSVEIKIRPRGTTVTVEPEKLSFRVVGOKLSFVRAVKTVEV
KLSFGPATVETGHIWNSDGRNVTSLVLTLOQL"
complement(join(10008..10128,10218..11056))
/notes="emb|CAB67676.1
gene id: MLN21.3
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01031.1"
/db_xref="GI:9279573"
/translations="MGSSSSSSSTKVDTMLVDETVLSALFDYDFIIPFADSEYATEI
HLOEAMFSGFAGTVGNHPOVHRNLVTLVQEPKIEAENFMPSPRLCMICMDEK
PSSDIIFRGNTYCTHACTVTRYVATIKENASRIKCPDVECTRLIEPTVCRDLIPK
DVPDRWEKILCESLISWDKFCYCPKDCSAMVNNENGANDVNTQECSCHELFVQCQ
KVTWHAGICDDEQRGNTKKSSDEDDALLQMAKAKQWRRCPCCKFVYDKVEGQCH
IKRISVLWMLTICVFFGRMENSYLKDGQSMYSINLFRKYVEF"
complement(join(12084..12461,12563..12877))
/notes="gb|AA02862.1
gene id: MLN21.4
similar to unknown protein"

```

CDS

```

/evidence=not_experimental
/protein_id="BAB01033.1"
/db_xref="GI:9279575"
/translations="MGTRDSNRTFSEIVGLIKSWLPWRGEPATVSRDFWMPQSCRV
CYECDCOFTLINRRHCHGRCVFCGCKTANSIPFAPSDLRTREDWELRVNVCYCFR
QWEGCDGPHVSNITELSTSPSTSLSSKSTSTANSSSFALGMPGLIGLGNRVHGG
DSVLHGSGNEMTSVTKQKRETSRFSFIATDVEDPSFALNSIRSDDEYDEYAYOT
GIEFTSHPRANEYYPGMEYNGMIDDPVCHLGETADQKLSGLSPHILGDSLELIRE
DSQFQKSHGDRDECEASPADISDDQVVEPVFENGLLWVPPEDEBERESA
LFDENEDGASGEGYLRPSTSGSGYGEDRTTEHKAVNKVVDGHFRALLAOL
LOVENISVSDDEGKESWLEIITSLWEAANLLKDMKSGMDPGGVYKVKCLASGR
HDSNVKGVCKKNVNRMSSTIKERALLILGGLEYQRVSNQLSDFLLQOEKDH
LKVDAKIHAEFENILLVEKSVSRFAEQEYLLAKDISLVNLIKRPDLLRIARCTAQII
PSVDHLSSQKLYGCENFRVDYPEEHGSTQGVKVKVTKLMYFHECPKPLGFTILLRG
ANEDELKKVHVQVGYFAAYHLALETSLFADGASPELNPSTIVALPDKSTIER
SISTVPGFTVSTVEKSPTMLSCAEPORANSVPVSELLSTTNLSIQNDIPIYVGSWG
QAREINSPFVSRHINISLNLDPVIESRNSDLSRSVPVDTADKSNPIVADETTNN
SLHUGGQFVRKSGQIGTSIWENQDNGSELTIAQQNNKPKETQCKEFPFSPSD
HQSLTSLRSRSMVKGTVCERSHLRIKYKYSFDPKPLGRFLDHLFDQSVCRSCEMP
SEAHVCHTRQGLSLTISVKQLDYLPGKEGKIKWMHRCLEPRLNGPFPATLRIV
MSDAWMLSTGKFLLEISFSNHAASRVACGSHLSHRDCLAFYCGNMAFCRATIVV
HSVLPSPSILSFENVENODWICQRTDEVIERAEALLFSEVLNAISQIAEKGRFRRIE
EVLQKAEFEENMQKILHREVNEGQPLVDILELYRIHQQLPQSYWMDHRLINASTL
HKUNSDOTKRENEKPLAKSOTLPEMNAGTNSLUGSEVNLNPDGSDTGDTCGLNN
VQEAADTNSLYQEKDDGVSFSPKLPDTSYPLENKVDYRTQSDQIYMKNLSATL
DAAMIGRQTSVEIPTNNKVSPLPSTMSNSTFPPISEGLMPIDL PQQNEFKVAYPV
SPALPSKNVSEDSVSWLSPVPLNFVRSINKNFISSOKLDTFGEHSPIVSSFREA
ELOGQPRLLLPVGLNDIVPYDDEPFSMIAYALMSPEYQRTSAEGESLVSYSELN
IPRPVDTTIIDFPRSNGSVDESILSSSSSTSLDDPLSYTKALHARVSYGEDGTGK
VKYTVTCYAKRFEALRGICLPESEYIRLSRCKWGAQSGKSNVFAKTLDDRTFI
KQVTKLEBSFIKPAFAYFKLSEISTKPTCLAKILGIYQVATKQKSGKTKMDV
LIMENLFGTIVKRLVDLKSRRARYNPDSGSKNLLDONLTEAMPTSPIFVGNKAK
RILERAVNDTAFALGDVMDYSLVGVDEKNELVGIIDFLRQYTWDXHLSWVKEF
TGILGGRNAPTVISPKQYKRRFRKAMTYFLMVPDQWFPNVVANNSSKDQPEETS
QAGTQAE"
complement(join(25752..25887,25980..26104,26220..26384))
/notes="gene id: MLN21.7
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01034.1"
/db_xref="GI:9279576"
/translations="MAYSTKYLPLKTPCFVPKTVAKTLMLOQSOVHVVKVATVSO
POPRTYKQSLVREDPIQGIICVTDNNGEVICEYDGGPCPTQSPWVASYREVEL
DLLOSQKSLSDSKQGRFIIASEQFMMKSSDFFL"
join(28462..28475,28608..28689,28882..28962,29255..29368,
29463..29564,29876..29934,30031..30133,30228..30293,
30387..30479)
/notes="gene id: MLN21.8"
/codon_start=1

```

CDS

```

/evidence=not_experimental
/protein_id="BAB01035.1"
/db_xref="GI:9279577"
/translations="MAGFVFFLLTSSPSSASSNSLTVIVHVDHEAKPSIFPTH
HTYTSLSASLTSPSIIHTYDVFHFSARLTSQDASQLLDHPHVSIVPEQVRLH
FHTSPFEGLRSTQAGLLESDFDGLVIGDVTGWPERPSPDRLGLGPVPIKWK
GQICASQDPESACNKLKVGARFCGGYEATNGKNETTEFRSPRSDGHHGHTASIS
AGRYVPASTLGVAHGVAAGMAPKARLAKYKVCWNSGCCYDSILAPATVADGVDVI
SLSVGVVVPYIDAIFAGFAIDRGIFFVSASAGNGPCALTITNAPWMTTVGAGT
IDRFANVKGKNGKMSIGSVYVGGPLDPRMYPLVYGGSLGDEGSSSLCEGLS
DPLNVKGKIVLDRGINSRATKEIVRKNGLGMIITANGFDGGLVADCHVLPATSV
GASGGBEIRRYIESKSRSSKHTATIVFKGRLGIRPAPVAVGSFARGPNPTEPI
LKPDVITAPGNILAAWPDRIGPSVTSNRTTEENTLSTSNACHVSLAALLAAH
PDWSPAIRSAITATYVDNNGEPMWDESTNGTSSVMDYGSGHVHTKAMPDGLVID
ITSYDINFLCNSYTRTNVITIRQACDCCARRAGHVGYNLNPFSFVYFOQGESK
MSTHTRITVNGDSVSVEIKIRPRGTTVTVEPEKLSFRVVGOKLSFVRAVKTVEV
KLSFGPATVETGHIWNSDGRNVTSLVLTLOQL"
complement(join(10008..10128,10218..11056))
/notes="emb|CAB67676.1
gene id: MLN21.3
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01031.1"
/db_xref="GI:9279573"
/translations="MGSSSSSSSTKVDTMLVDETVLSALFDYDFIIPFADSEYATEI
HLOEAMFSGFAGTVGNHPOVHRNLVTLVQEPKIEAENFMPSPRLCMICMDEK
PSSDIIFRGNTYCTHACTVTRYVATIKENASRIKCPDVECTRLIEPTVCRDLIPK
DVPDRWEKILCESLISWDKFCYCPKDCSAMVNNENGANDVNTQECSCHELFVQCQ
KVTWHAGICDDEQRGNTKKSSDEDDALLQMAKAKQWRRCPCCKFVYDKVEGQCH
IKRISVLWMLTICVFFGRMENSYLKDGQSMYSINLFRKYVEF"
complement(join(12084..12461,12563..12877))
/notes="gb|AA02862.1
gene id: MLN21.4
similar to unknown protein"

```

CDS

```

/evidence=not_experimental
/protein_id="BAB01034.1"
/db_xref="GI:9279576"
/translations="MAYSTKYLPLKTPCFVPKTVAKTLMLOQSOVHVVKVATVSO
POPRTYKQSLVREDPIQGIICVTDNNGEVICEYDGGPCPTQSPWVASYREVEL
DLLOSQKSLSDSKQGRFIIASEQFMMKSSDFFL"
join(28462..28475,28608..28689,28882..28962,29255..29368,
29463..29564,29876..29934,30031..30133,30228..30293,
30387..30479)
/notes="gene id: MLN21.8"
/codon_start=1

```

CDS

```

/evidence=not_experimental
/protein_id="BAB01034.1"
/db_xref="GI:9279576"
/translations="MAYSTKYLPLKTPCFVPKTVAKTLMLOQSOVHVVKVATVSO
POPRTYKQSLVREDPIQGIICVTDNNGEVICEYDGGPCPTQSPWVASYREVEL
DLLOSQKSLSDSKQGRFIIASEQFMMKSSDFFL"
join(28462..28475,28608..28689,28882..28962,29255..29368,
29463..29564,29876..29934,30031..30133,30228..30293,
30387..30479)
/notes="gene id: MLN21.8"
/codon_start=1

```

```

/codon_start=1
/evidence=not_experimental
/protein_id="BAB01032.1"
/db_xref="GI:9279574"
/translations="MQYRRPFETKTQYRICPAETMVKIHDPDLTISGAGEBTSSPYL
TTEOSFTIWMKSLVFNCTGTFDSKGNIIYRVDNVNSCREVILMDLSGHVLFLL
RQKFGELPKTWEGYRSSSATVETTKLEYFRVKNVQVQPNKQSSSSYVNASCRND
EQCYKRWTVTSGSSLAIEDNCKLAEVKRKQSLNGLKDDVDVTMMVSESQVDHRSFIIG
LVLAHSLINCIL"
complement(16501..16918)
/notes="gb|AAB63539.1
gene id: MLN21.5
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental
join(18722..19329,19578..20327,20473..20626,20740..21084,
21193..22037,22134..22465,22571..22694,22775..22903,
23041..24525,24603..24869,24957..25283)
/notes="emb|CAB36798.1
gene id: MLN21.6
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01033.1"
/db_xref="GI:9279575"
/translations="MGTRDSNRTFSEIVGLIKSWLPWRGEPATVSRDFWMPQSCRV
CYECDCOFTLINRRHCHGRCVFCGCKTANSIPFAPSDLRTREDWELRVNVCYCFR
QWEGCDGPHVSNITELSTSPSTSLSSKSTSTANSSSFALGMPGLIGLGNRVHGG
DSVLHGSGNEMTSVTKQKRETSRFSFIATDVEDPSFALNSIRSDDEYDEYAYOT
GIEFTSHPRANEYYPGMEYNGMIDDPVCHLGETADQKLSGLSPHILGDSLELIRE
DSQFQKSHGDRDECEASPADISDDQVVEPVFENGLLWVPPEDEBERESA
LFDENEDGASGEGYLRPSTSGSGYGEDRTTEHKAVNKVVDGHFRALLAOL
LOVENISVSDDEGKESWLEIITSLWEAANLLKDMKSGMDPGGVYKVKCLASGR
HDSNVKGVCKKNVNRMSSTIKERALLILGGLEYQRVSNQLSDFLLQOEKDH
LKVDAKIHAEFENILLVEKSVSRFAEQEYLLAKDISLVNLIKRPDLLRIARCTAQII
PSVDHLSSQKLYGCENFRVDYPEEHGSTQGVKVKVTKLMYFHECPKPLGFTILLRG
ANEDELKKVHVQVGYFAAYHLALETSLFADGASPELNPSTIVALPDKSTIER
SISTVPGFTVSTVEKSPTMLSCAEPORANSVPVSELLSTTNLSIQNDIPIYVGSWG
QAREINSPFVSRHINISLNLDPVIESRNSDLSRSVPVDTADKSNPIVADETTNN
SLHUGGQFVRKSGQIGTSIWENQDNGSELTIAQQNNKPKETQCKEFPFSPSD
HQSLTSLRSRSMVKGTVCERSHLRIKYKYSFDPKPLGRFLDHLFDQSVCRSCEMP
SEAHVCHTRQGLSLTISVKQLDYLPGKEGKIKWMHRCLEPRLNGPFPATLRIV
MSDAWMLSTGKFLLEISFSNHAASRVACGSHLSHRDCLAFYCGNMAFCRATIVV
HSVLPSPSILSFENVENODWICQRTDEVIERAEALLFSEVLNAISQIAEKGRFRRIE
EVLQKAEFEENMQKILHREVNEGQPLVDILELYRIHQQLPQSYWMDHRLINASTL
HKUNSDOTKRENEKPLAKSOTLPEMNAGTNSLUGSEVNLNPDGSDTGDTCGLNN
VQEAADTNSLYQEKDDGVSFSPKLPDTSYPLENKVDYRTQSDQIYMKNLSATL
DAAMIGRQTSVEIPTNNKVSPLPSTMSNSTFPPISEGLMPIDL PQQNEFKVAYPV
SPALPSKNVSEDSVSWLSPVPLNFVRSINKNFISSOKLDTFGEHSPIVSSFREA
ELOGQPRLLLPVGLNDIVPYDDEPFSMIAYALMSPEYQRTSAEGESLVSYSELN
IPRPVDTTIIDFPRSNGSVDESILSSSSSTSLDDPLSYTKALHARVSYGEDGTGK
VKYTVTCYAKRFEALRGICLPESEYIRLSRCKWGAQSGKSNVFAKTLDDRTFI
KQVTKLEBSFIKPAFAYFKLSEISTKPTCLAKILGIYQVATKQKSGKTKMDV
LIMENLFGTIVKRLVDLKSRRARYNPDSGSKNLLDONLTEAMPTSPIFVGNKAK
RILERAVNDTAFALGDVMDYSLVGVDEKNELVGIIDFLRQYTWDXHLSWVKEF
TGILGGRNAPTVISPKQYKRRFRKAMTYFLMVPDQWFPNVVANNSSKDQPEETS
QAGTQAE"
complement(join(25752..25887,25980..26104,26220..26384))
/notes="gene id: MLN21.7
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01034.1"
/db_xref="GI:9279576"
/translations="MAYSTKYLPLKTPCFVPKTVAKTLMLOQSOVHVVKVATVSO
POPRTYKQSLVREDPIQGIICVTDNNGEVICEYDGGPCPTQSPWVASYREVEL
DLLOSQKSLSDSKQGRFIIASEQFMMKSSDFFL"
join(28462..28475,28608..28689,28882..28962,29255..29368,
29463..29564,29876..29934,30031..30133,30228..30293,
30387..30479)
/notes="gene id: MLN21.8"
/codon_start=1

```



```

/evidence=not_experimental
/product="20S proteasome subunit PAE-like protein"
/protein_id="BAB01035.1"
/db_xref="GI:9279577"
/translation="MLTRTEYDRGVNTSPESGRIFQVEYVAIEAIKLGSTAIGVKMKE
GWLVAEKRTISPLIPSSVEKMEIDDHIGCMSGLIADARTLVEHARVETQNHKPS
YGEWTVETSTQALCDLAKRFGEESERFVGSLLIAGHDENGSLFTDPSGTF
WQCNAKAGSGSGASQSSQEQNKDITLQEAETLAVSLIKVMKEEKYTPNNVDIAKV
APAYHLTYPOEVZAVTSRL"
CDS
complement(join(31003..31685,31773..33996))
/note="gene_id:MLN21.9"
/codon_start=1
/evidence=not_experimental
/product="pectinesterase-like protein"

Alignment Scores:
Pred. No.: 2,64e-39 Length: 83906
Score: 1700.00 Matches: 356
Percent Similarity: 61.95% Conservative: 134
Best Local Similarity: 45.01% Mismatches: 237
Query Match: 42.31% Indels: 44
DB: 8 Gaps: 15

```

US-09-806-767-2 (1-775) X AB022220 (1-83906)

[illegible]

Qy	243	AsnGlyAlaGlyValAlaAArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysVal	262
Db	7372	TAGCCTCACGGCGTCGCTGCTGGAGTAGGCTCCAAAAGCTAGACTCGCGCTTCAAAGTC	7313
Qy	263	CysTrpPheAsnGlyCysTrpSerSerAspIleLeuAlaIleAspValAlaIleGln	282
Db	7312	TGTTGGAACTCCGGTGTACGACTCGATATCTTAGCCGCTTCGACACCGCGGTGCC	7253
Qy	283	AspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAsp	302
Db	7252	GACGGTGCATGTCTCTCTCTCGTTGGAGGCGTCGTGCTTCTATTACCTAGAC	7193
Qy	303	ThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAla	322
Db	7192	GCTATCGCTATTGGAGCTTTTGGAGCTATTTCAGAGAGGCATATTCGCTCTCTGCTCCGCC	7133
Qy	323	GlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIle	342
Db	7132	GGTAACGGAGTCCGGTGTCTTAAACGGTGACAAATGTTGCTCCGTGGATGACACAGTC	7073
Qy	343	GlyAlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeu	362
Db	7072	GGAGCTGGACAATCATAGGATTTTCGGGCCAATCTGAAATCTGGTAACGGGAAGATC	7013
Qy	363	LeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGlu	382
Db	7012	ATTCTCGTGTAGTGATATACGGTGGACCGGGCTCTG---GATCGGGTCAATGTATCCCG	6956
Qy	383	ValIleTyr-----ValThrGlyGlyAsp---LysGlySerGluPheCysLeuArg	398
Db	6955	CTTGTTTACCGGTGGTAGTTTACTAGGCGCGGATGGTTCATCTGCTCTGTGTCCTTGAA	6896
Qy	399	GlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArgGlyValAsn	418
Db	6895	GGCTCGTGTGATCCGAATTTAGTGAAGGAAGATCGTCTCTTTGTGATAGAGNAATCAAT	6836
Qy	419	GlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAla	438
Db	6835	TCTAGAGCAACCAAGGTGAGATCGTACGCAAAATGAGGCTTGGGGATGATTATAGCG	6776
Qy	439	AsnThrCluIleAsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeu	458
Db	6775	AATGGTGTTGACCGGTGAAGTTTATGATGCTGATTCACCGCTTTACCGGACATCT	6716
Qy	459	IleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLys----	476
Db	6715	GTGGTGCTCTCGGAGGAGTGAATCTGATGATATCTCTGAATCATCCAAATCTCGT	6656
Qy	477	-----ProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArg	492
Db	6655	TCATCGAAACATCCACGGCTACGATGTGTTTCAAAGGTACCCGACTGGGATCCGACCT	6596
Qy	493	AlaProGluValAlaGlnPheSerAlaArgGlyProSerSerLeuAlaAsnProSerIleLeu	512
Db	6595	GCACCGGTTGTGCATCTTTCTCTGCTCGTGGTCTTAATCCAGAGACGCCGCGAGATTCCT	6536
Qy	513	LysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGly	532
Db	6535	AAACCGGATGTAATCCGACCTGGTTGCAATATTTCTAGCTGCTTGGCTGACCGGATTTGGT	6476
Qy	533	ProThrGlyLeuProTyrAspSerArgValAsnPheThrValMetSerGlyThrSer	552
Db	6475	CCATCTGGTGTACTTCTGATACCCGGAGAACTGAGTTCAAACATTTTATCAGGCACCTCG	6416
Qy	553	MetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrp	572
Db	6415	ATCGCGTCCCGCACGCTGCTGCTAGCTGCTTTGCTTAAAGCGCGCTCATCCGGATTTGG	6356
Qy	573	SerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGly	592
Db	6355	AGTCCGCGCGCATAGATCAGCATTTGATAACCAACCGCTTACACGGTTGATAACACGGGT	6296
Qy	593	LysAlaIleLysAsp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAla	609

/notes="synonym: T6P5.4; identical to GB:AF043519; supported by cDNA: gi.15450622.gb.AV052679.1" join(18080..18214,18297..18393,18473..18554,18619..18715,18791..18849,18932..19014,19321..19416,19606..19665,19874..20180) /gene="At2g05840" join(18139..18214,18297..18391,18473..18554,18619..18715,18791..18849,18932..19014,19321..19416,19606..19665,19874..19966) /gene="At2g05840" /codon_start=1 /product="20S proteasome alpha subunit A2 (PAA2)" /protein_id="AAC95161.1" /db_xref="GI:4006820" /translation="MSRGGAGYDRHITIFPEGRLFQVEYAFKAVKAGITSGVRG KSVCVVTKQKVPDKLLDQSSVHLFPVTKYLGILGATGNTADRSLSVQARNEAEFR FQYGYMPADIKAHTADKSQVYTHAYMRPLGVAMVLGIDDEERGPLYKCDPAGHF YGKATSAAGKQEAQVNFLEKMKENPAFTYDTVTQTALQSALQVLDKFRATEIEVG VVADDPFRSLRTEIDELHTAISER" complement(23222..23330) /rpt_family="AT_rich" complement(23980..24022) /rpt_family="AT_rich" complement(24550..24574) /rpt_family="AT_rich" complement(24766..24786) /rpt_family="AT_rich" complement(24944..25022) /rpt_family="AT_rich" 25297..25317 /rpt_family="AT_rich" 25827..25856 /rpt_family="AT_rich" complement(25903..25966) /rpt_family="AT_rich" complement(26687..28373) /gene="At2g05850" /notes="synonym: T6P5.5" complement(join(<26687..27696,27802..27909,28028..>28373)) /gene="At2g05850" complement(join(26687..27696,27802..27909,28028..28373)) /codon_start=1 /product="putative serine carboxypeptidase II" /protein_id="AAC95162.1" /db_xref="GI:4006820" /translation="MCKQDWSVTACIFLSLSIASIHCSQTHPPSHKGGAGLSGDT SHFNSVRNVLSEKDLIELKLPGPSGISPROYGVVAVNEPATRFLVYFVFAIK PSKSTFLVNFNGPCSSVGGAPFELGPRVHSDCKTLNRPYSWNNEANMLFFEG PISVGSFSTPPDWEIFGEQADKLTAEINDMPLVNWLERPEYKGRDVIYSGQSYAG HYTPQAQILHNNQTFINLRGISGNFGLDLLEADNENKFIHLGSLVSKQFEY SKYCDPANDMDPCPKIPKPSIEHNKHLVDVNIYAPVCLNSTLSPEPKCTTMEVD PCSNVYKALVNSENVQEAHANTKLPYEWKACHYLSNVWDADKQASMPILHDL MCGSRVLYVSGVDVAALPFTATMAVLTKMLITVNWPRPWFETGGQLGFTEDYENL TYARVKGSHSVLPDQFVALNLFSTFIRNLPQTP" 28927..28981 /rpt_family="AT_rich" complement(29152..29229) /rpt_family="AT_rich" complement(31013..32021) /gene="At2g05860" /notes="synonym: T6P5.6" complement(join(<31013..31223,31303..31405,31542..31806,31884..>32021)) /gene="At2g05860" complement(join(31013..31223,31303..31405,31542..31806,31884..32021)) /product="putative replication protein A1" /protein_id="AAC95163.1" /db_xref="GI:4006821" /translation="MESFHLKSLKSRGWCIRGVRWTFLLVSLVSPSKVMGLILAD

Alignment Scores:

Pred. No.:	6.34e-99	Length:	121160
Score:	1697.50	Matches:	370
Percent Similarity:	60.47%	Conservative:	115
Best Local Similarity:	46.13%	Mismatches:	241
Query Match:	42.25%	Indels:	76
DB:	8	Gaps:	16

US-09-806-767-2 (1-775) x AC005970 (1-121160)

QY	11	IlePheLeuLeuPheCysSerSerSerGluLeuLeuGlnLysGlnThrThrIleVal	30
DB	56069	CTCTCTCTCTCTCTCCACACACCGCC-----AAAAAACCCTCATCATC	56025
QY	31	GlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTTPHisLeu	50
DB	56024	CGCGTCAAT---CACTCCGATAAACCGGAATCATTCCTCACTCACGAGTTGGTACACA	55968
QY	51	SerPheLeuGlnGluAlaValLeuGlyValIleGluGluGluGluProSerSerArg	70
DB	55967	TCTCAACTCAATCA-----GAATCATCT	55944
QY	71	LeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGlu	90
DB	55943	CTTCTCTACACTTACACACACTCTCTCCATGGCTTCTCCGCTTACCTCGACTCCGCGAA	55884
QY	91	AlaGlu---IleLeuArgTyrSerProGluValValAlaValArgProAspHisValLeu	109
DB	55883	GCCGATTCTCTCTCTCTCTCTCAAACTCAATCTCGATATCTTCGAAGATCTCTCTAC	55824
QY	110	GlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGly---PheGlyAsnSer	128
DB	55823	ACATTTCACACTACGCGTACTCTCGAGTTCTCGGTTCTCAATCCGAAATTCGGTGTTCAC	55764
QY	129	GlyValTrpSerTyrSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGly	148
DB	55763	GATCTCGGTTCTTCTTCT-----AACGGGCTTATCATCGAGTTTATAGATACTGGC	55713
QY	149	ValTrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrp	168
DB	55712	GTTTGGCGCTGAATAGAACCTCGATGATGATGATGATGATGATGATGATGATGATGATG	55653
QY	169	LysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLysLeuIle	188
DB	55652	AAAGAGAATGTGAATCTGGTTCGATTTCGATTCCAAAGTTGTATAAAGAGCTTATC	55593
QY	189	GlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsn	208
DB	55592	GGACTAGAGCTTCTCCAAAGGATTTCAATGGCTTCTGGTGGTGGTGGTGGTGGTGGTGG	55533
QY	209	MetProArgGluThrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSer	228
DB	55532	-----CGTGAATCTGTTTCTCTCGTGTGATGTTTCACGACATGGAACACATCTTCACT	55479
QY	229	ThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAla	248
DB	55478	ACCCCGCGGATCCCGGTTAGAAACGCTAGTCTCTCGGTTACGCCCGGTTACGCC	55419
QY	249	ArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys	268
DB	55418	AGAGGTATGGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	55359
QY	269	TyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeu	288
DB	55358	TTTGGATCTGATATCTAGCTGCTATGATCGAGCTATCTTGTGATGCTGCTGCTGCTG	55299
QY	289	SerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThr	308

Db 55298 TCGTTATCTCTGGTGGTCTGCTCCGATATATTCGCGATACGATTCGATTCGATTCGAGCG 55239
Qy 309 PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle 328
Db 55238 TTTTCGGCTATGGAGAGGTGTTTGTCTCTGTTCTCTGTAATAGTGGTCTACT 55179
Qy 329 GluSerSerValAlaAsnThrAlaProTIPValSerThrIleGlyAlaGlyThrLeuAsp 348
Db 55178 AGAGCTTCTGTGCAATGTTGCTCTGCTGTTGATGACTGTTGGTCTGCTGTTAGAT 55119
Qy 349 ArgArgPheProAlaValAlaArgLeuAlaAsnGlyLysLeuLeuTyrglyGluSerLeu 368
Db 55118 AGAGATTTTCGGGTTTTTCGCAATCTCGTAACGGGAACGACTTACCGGTGTTTCGCTG 55059
Qy 369 TyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyValThrGly 388
Db 55058 TATAGCGGTAGGAATG-----GGACGAAGCGGCTTGAATGGTTTATATAAAGG 55005
Qy 389 GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGly 408
Db 55004 AATAGTAGTTCGAGTAATCTTTGTTTACCTGGTTGCTTCGATTTCGATGATTGCTGGG 54945
Qy 409 LysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLys 428
Db 54944 AAGATTGTTGTTGATAGAGTGTAAATGCTAGAGTTGAGAAGAGAGTGTGTGTAG 54885
Qy 429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrCluIleAsnGlnGluAspSer 448
Db 54884 GATCTGGTGGTTTAGGAGTGAATGCGGAATACGCTCGGAGTGGAGAGGAGCTTGTG 54825
Qy 449 IleAspValHisLeuLeuProAlaThrLeuIleGlyTyThrGluSerValLeuLys 468
Db 54824 CGCGATAGTCAATTGCTTCGCGGATCGCTGTAGGAGAGAGACTGGTATTACTTAGG 54765
Qy 469 AlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyThrValIle 488
Db 54764 GAGTATGTTAAGTCAGATCTAAACCAACCGCTCTCTGTTTAAAGGAACGGTCTT 54705
Qy 489 GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsn 508
Db 54704 GACGTTAAGCGCTCTCTGTTGCTGCTTTAGCTCGAGAGGCTCTAATACTGTTACT 54645
Qy 509 ProSerIleLeuLysProAspMetIleAlaProGlyValAlaSerIleAlaIleAlaTTPPro 528
Db 54644 CCGAATCTTGAAGCTGATGTTATTTGCTGAGGATTAATATTTGGCTGGTGGTCT 54585
Qy 529 GlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgArgValAsnPheThrValMet 548
Db 54584 GACGCTATTGGTCTCTACTGTTCTTGACAAGGACTCTAGAGGACTCAGTTCAACATCG 54525
Qy 549 Ser----- 549
Db 54524 TCAGGTACTCAACTTGAACTTGATTTGATTACATATTTTGTATTGATTATATCTAT 54465
Qy 550 -----Glyth 551
Db 54464 GATTGTAGAGAGGGGTTTGTGTTTGTGACATGAAAGTCTCGATTATTTGTTAGGTAC 54405
Qy 551 rSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAs 571
Db 54404 GTCATGTCAAGCCCAACATCAGTCAGTGGTTAGCGGGTCTTTTGAAGCAGCTCACCTGA 54345
Qy 571 nTTPSerProAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGly 591
Db 54344 GTGGAGTCCGAGTCTATCAATCAGCTCTCATGACTACAGCTTACGTTCTTCGACAACAC 54285
Qy 591 nGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAlaIleGlyAl 609
Db 54284 CAACGCTCTCTCCATGATGTCGACAGCAACAGCCTATCTAACCCATATCTCTCACGGCTC 54225
Qy 609 aGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProVa 629
Db 54224 GGGCCATGTAGATCCCCAAAAGGCTCTCTCACAGGCTTGTCTACGACATCTCAACCGGA 54165

Qy 629 lAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleTh 649
Db 54164 GGAATACATCAGGTTTTTGTCTCTAGACTACACAGTCGATCAGATTGTTGGATTGT 54105
Qy 649 rHisLys---AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAs 668
Db 54104 GAAGCGACTACGCTTAACTCTCGAAGAAGTTC---TCAGATCTGCT---CAGCTCAA 54051
Qy 668 nTyProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArgAr 688
Db 54050 CTACCCAGTTTTCGGTTTTTGT---GGGGTAAAGAGAGTTGTCGGTACATCTCGGA 53994
Qy 688 gValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGl 708
Db 53993 AGTAACAATCTGGTGCAGCAAGCTCGTTTACAAAGTGACGGTTAATGGAGCTCCTAG 53934
Qy 708 yIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSe 728
Db 53933 TGTGGAAATCTCTGTTAAACCATCGAAACTTTCGTTTAAAGCGTGGGAGAGAAGAG 53874
Qy 728 rTyArgValTrpPheValLeuLysLysAsnArgGlyLysValAlaIleSerPheAl 748
Db 53873 GTACACAGTCACTGTTTACCAAGAAAGGAGTGAGTATGACGAACAGGCTGAGTTT-- 53816
Qy 748 aGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProil 768
Db 53815 ---GGTTCATCACTTGGACCAATCCGACAC-----GAAGTGAGAGATCCCGT 53769
Qy 768 eSer 769
Db 53768 TGCA 53765

RESULT 10
AY084387
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AY084387 2638 bp mRNA linear PLN 27-MAR-2003
Arabidopsis thaliana clone 106539 mRNA, complete sequence.
AY084387
AY084387.1 GI:21403097
FLI:CDNA.
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 2638)
Haas, B.J., Volfovsky, N., Town, C.D., Troughan, M., Alexandrov, N.,
Feldmann, K.A., Flavell, R.B., White, O. and Salzberg, S.L.
Full-length messenger RNA sequences greatly improve genome
annotation
Genome Biol. 3 (6), RESEARCH0029 (2002)
2 (bases 1 to 2638)
Brover, V., Troughan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and
Feldmann, K.
Full-length cDNA from Arabidopsis thaliana
Unpublished
3 (bases 1 to 2638)
Brover, V., Troughan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and
Feldmann, K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A


```

Db 1621 GCACCGGTTGTCATCTTCTCTGACGTGTCCTAATCCAGACGCGGAGATCTT 1680
Qy 513 LysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTTPProGlnAsnLeuGly 532
Db 1681 AAACCGGATGTAATCCACTGCTTGAATATCTAGCTGCTGGCCTGACCGGATGGT 1740
Qy 533 ProThrGlyLeuProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSer 552
Db 1741 CCATCTGGTGTACTTCTGTAACCGGAGAACTGAGTTCAACATTTTATCAGGCACCTCG 1800
Qy 553 MetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrp 572
Db 1801 ATGGCGTGGCCCATGTTCTGCTGCTAGCTGCTTGTCTTAAGCGGCTCATCCGATGG 1860
Qy 573 SerProAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGly 592
Db 1861 AGCCCGCGCGGATAAGATCAGCATTTGATGACACGGCTTACACGGTTGATAACACGGT 1920
Qy 593 LysAlaIleLysAsp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAla 609
Db 1921 GAGCGATGATGATGAGTCCACTGCCAT---ACGCTCTCGTTACGGATATGCTTGG 1977
Qy 610 GlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProVal 629
Db 1978 GGTGATGTTCAACCAACAGAGCATGATGATCGGGACTAGTCTACGATATAACATCTTAT 2037
Qy 630 AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThr 649
Db 2038 GATTATATCAACTTCTGTGTAATCTCACTACATCGGAACCAACATTTGTGACGATAACC 2097
Qy 650 HisLysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyr 669
Db 2098 GTGCGCAAGCGGACTGCGAGCGGTGCGAGCGGCGGACACGTCGGAATTTGAACATC 2157
Qy 670 ProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThr----- 686
Db 2158 CCGTCTCTCTGCTGCTGTTTCAGCAGTATGAGAGAGTAATAATGTCGACATTTTCAT 2217
Qy 687 ArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaPro 706
Db 2218 AGGACTGTGACAAATAGACGATTCGGATTCGGTTTACGATTAAGATTAGGCGGCT 2277
Qy 707 GluGlyLysValIleValAsnProLysArgLeuValPheLysHisValAspGlnThr 726
Db 2278 AGAGGACTACATGACCGTGTAGCGGAGAGCCTATCCTTCAGACGGTGGGCGAGAA 2337
Qy 727 LeuSerTyrArgValThrPheValLeuLysLysAsnArgGlyGlyLysValAlaSer 746
Db 2338 CTGAGT-----TTTGTGTTAGGTTAAGACACAGAGGTGAGTTGTACCA 2385
Qy 747 PheAla-----GlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGln 762
Db 2386 GGAGCGACAAATGTGAGAGCTGTATATGTTTGTGTCAGATGG-----AAACGG 2436
Qy 763 ArgValArgSerProLysValThrLys 773
Db 2437 AATGTGACAAAGTCCCTTGGTGTCTACTTCAA 2469

RESULT 11
LESBT2
LOCUS 2650 bp mRNA linear PLN 18-FEB-2000
DEFINITION L.esculentum mRNA for serine protease, SBT2.
ACCESSION X98930
VERSION X98930.1 GI:1771161
KEYWORDS SBT2 gene; serine protease.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
1 Janzik, I., Macheroux, P., Amrhein, N. and Schaller, A.

```

```

LeSBT1, a subtilase from tomato plants. Overexpression in insect
cells, purification, and characterization
J. Biol. Chem. 275 (7), 5193-5199 (2000)
MEDLINE 20138270
PUBMED 10671566
REFERENCE 2 (bases 1 to 2650)
AUTHORS Schaller/A.
DIRECT SUBMISSION
SUBMITTED (01-JUL-1996) A. Schaller, Federal Institute of
Technology, Institute of Plant Sciences, ETHZ, LFW E51,
Universitaetstrasse 2, CH-8092 Zuerich, SWITZERLAND
FEATURES
source
1..2650
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/sub_species="UC82b"
/db_xref="taxon:4081"
/clone_lib="Lamba ZAP"
/dev_stage="leaf"
1..2650
/genes="SBT2"
116..2443
/genes="SBT2"
/function="serine protease"
/codon_start=1
/evidence=not experimental
/protein_id="CAA67430.1"
/db_xref="GI:1771162"
/db_xref="SPTREMBL:P93205"
/translacion="MAGMLLKCMFFVSVCILAKSPNTKKTIIQMDKWAKPDV
FVHVOMYSLVKSLVPSVTEVEGTGDBERILYSYQATPHGVAQLSEEVKKLOER
NGVLAYPEIKYOLHTTRPFLGLDREDSSKLWADRLSDHNVIGVLDTGIPESPS
FNDGWTSPVSHMKVCETGRGFEKHCCKIIVARFVRGVEAASGKINERGFESA
RDQGHGTHGTAGVSVRGANLLGYATGAKMAPARVAAKYKWCWGGCRRSDIL
SAVDQVADGVNLSISGGVSSYNRSLIAAFAGMKEGVFVSCAGNGGDPDLSL
TNYSPWITTVGASTMDRPFATVETGKIVTGSALYKGMNLSLTKQYPLIYLGNS
SNLPSLCLDGLDKASVAGKIVICDRGISPRVQGVVKEAGVGMLITNAAAGE
ELVADSHLPAVAVGREGRAIKLYAAGRSATATLFLGTGKLRSPVVAFFSRGP
NFSLLEIKEDMVAQVNIAGWTGALGPSLIDQRRTNFNILSTSCSHVSGIA
ALAKARHPDMPAAIKSALMTIYVDNTIKSLKQASVTFPIYDHAGAHVNRKAV
EKLIDYDGAQDYFEFLCTQELSPQVMVFGKSNRTCHHSANPDGLNTPAISAVFP
DPTKLMLTLRTVTVGSPISNYHVVAFAKGVAVKVEPERLNFSTKNQKLSKYVTF
KTVSROKAPFEGSLWKDGTGTHKVRSPAIATWLASV"
BASE COUNT 756 a 511 c 618 g 765 t
ORIGIN
Alignment Scores:
Pred. No.: 5,45e-100 Length: 2650
Score: 1679.00 Matches: 359
Percent Similarity: 61.97% Conservative: 125
Best Local Similarity: 45.97% Mismatches: 261
Query Match: 41.79% Indels: 36
DB: 8 Gaps: 16
US-09-806-767-2 (1-775) x LESBT2 (1-2650)
Qy 6 PhePheLeuCysIle-----IlePheLeuLeuPheCysSerSerSerSerGluIle 22
Db 146 TTCTTTGATCAGTATGCTTGTCTTATCAACTACGCAAAATGTAGCCCAAACT- 199
Qy 23 LeuGlnLysGlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLys----- 40
Db 200 -----AAGAAACTTACATTATCCAAATG-----GATAAATGGGTAAACCAAGAT 244
Qy 41 ThrPheAlaSerLysPheAspTTPHLeuSerPheLeuGlnGluAlaValLeuGlyVal 60
Db 245 GTATTCGTCACCATGTTCAATGCTACTCATCTCGTAAATACAGTACTACCATCTACA 304
Qy 61 GluGluGluGluGlu-----GluProSerSerArgLeuLeuTyrSerTyrGlySerAla 78
Db 305 ACAGAGGTAGAAAAAACACGGGATGGAGAGAGGATACCTTTTACAGCTACCAAGCTGCT 364
Qy 79 IleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluLeuLeuArgTyrSerPro 98

```

Db 365 TTTTCATGGTGTTCAGCTCAGCTTGTAGTGAAGAGAAAGTGAAGAAATTTACAAAGACGAAT 424
 Qy 99 GluValValAlaValArgProAspHisValLeuGlnValGlnThrThrTyrSerTyrLys 118
 Db 425 GGGGTTTTCCTCTTCTTCTGAGATTAAGATACAGTTACACCTAGAGAGTCTCTG 484
 Qy 119 PheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTyrSerLysSerArgPheGlyGln 138
 Db 485 TTTCTTGGGCTTGACCGTGAGGATAGCAGTAAGTTGTGGCGGATAGATTATCTGACCAT 544
 Qy 139 GlyThrIleIleGlyValLeuAspThrGlyValTyrProGluSerProSerPheAspLys 158
 Db 545 AATGTAATTCGGGTGTCTTGTACTACTGGGATTTGGCGGAAAGTCCGAGTTTCAATGAT 604
 Qy 159 ThrGlyMetProSerIleProArgLysTyrLysGlyLysCysGlnIleGlyGluSerPhe 178
 Db 605 AGTGGATGACTAGTGTCTCATCTCTGGAAGGGGTTGTGAACAGGGAGGGGTTT 664
 Qy 179 SerSerSerCysAsnArgLysLeuIleGlyValArgPhePheIleArgGlyHisArg 198
 Db 665 GAGNAGCATCATTCGACGAGAAATTTGTGTGTGTAGTGTTTTCCGTGGGTATGAA 724
 Qy 199 ValAlaAsnSerProGluGluSerProAsnMetProArgGluTyrIleSerAlaArgAsp 218
 Db 725 GCAGCTTCTGCT-----AAATCAATGAAGGGGTGAGTTCAATCGGCAAGAGAT 775
 Qy 219 SerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAla 238
 Db 776 CAAGATGGGATGGTACTCATCTGCTGGAACAGTTGCGAGGGTCTGTAGTCCGTGGTCT 835
 Qy 239 AsnValLeuGlyAsnGlyAlaGlyValAlaAspGlyMetAlaProGlyValHisIleAla 258
 Db 836 AATCTTTTGGTGTATGCTTATGTACTGTCTAGAGGAATGCCCCCTGGTGGAGATTGCA 895
 Qy 259 ValTyrLysValCysThrPheAsnGlyCysTyrSerSerAspIleLeuAlaIleAsp 278
 Db 896 GCATATAAGTGTGTGGTGGGTGGATGTTTGTAGCTCAGATATATCTTCGAGTTGAT 955
 Qy 279 ValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyPheProIlePro 298
 Db 956 CAAGCTGTGCTGATGGGTTAATATCTGTCAATTTCTTGGGTGGTGGAGTTTCGTCT 1015
 Qy 299 LeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerVal 318
 Db 1016 TATAATCGTACAGTTGTCAATGTGCTTTTGGGGCCATGGAGAAAGGGTTTTGT 1075
 Qy 319 IleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTyr 338
 Db 1076 TCATGTTACAGCAAGAAATGTGGGCTGATCCAAATTAGTTGACAAATGTTCACCATGG 1135
 Qy 339 ValSerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAla 358
 Db 1136 ATCACTACTGTAGGAGCAACACTATGTAGTAGATTTTCCAGCACTGTTCGAGCTGGGG 1195
 Qy 359 AsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAlaGly 378
 Db 1196 ACGGGGAAGATTGTGACCGGAGCTTCACTTTATAAAGTAGTAGTAATCTTCAACACAA 1255
 Qy 379 ArgGluValGluValIleTyrValThrGlyGlyAspLysGly-----SerGluPhe 395
 Db 1256 AATCAGATTCCTTTGATTTTACTTAGGAAGTAATTCGAGCAATCTCATGCCGAGCTCATTTG 1315
 Qy 396 CysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArg 415
 Db 1316 TGCTTAGATGTTACCTTAGATAAGCTTCAGTTGCTGGAAATTTGTGATATGTGACAGG 1375
 Qy 416 GlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyValAlaMet 435
 Db 1376 GGAATTAGTCTCGAGTTCAAAAGGGGCAAGTTGTAAAGAGAGCTGGAGGATGAGAAATG 1435
 Qy 436 IleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeuLeuPro 455

Db 1436 ATTTTGACAAACACTCGACGCTAATGTGTGAAGAACTGGTGTCTGATTTCTCACCTTCTTCCG 1495
 Qy 456 AlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrVal 475
 Db 1496 GCAGTACAGTTGGTGAAGAGAGAGCAATCAAGCTCTACGCG---GCAGGTAGA 1552
 Qy 476 LysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
 Db 1553 AGCCTACTCCACTTTCAGATTTCCTGGGACTTAATTAGGATCAGGCTTCGCCCGTG 1612
 Qy 496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515
 Db 1613 GTTCTCCTCATTTTCATCCAGAGCACTAATCTCTCAGCTAGAGATTTCTAAAGCCAGAT 1672
 Qy 516 MetIleAlaProGlyValAsnIleAlaAlaTyrProGlnAsnLeuGlyProThrGly 535
 Db 1673 ATGTGTGCACAGAGTGAATATCTTCTGCTGTGACTGGTGCATGGTCCGTCTAGT 1732
 Qy 536 LeuProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCys 555
 Db 1733 TTGCCAATTGATCAACGAGGACGAATTTCAATATATTTGTCTGGAATTTCAATGTCTGT 1792
 Qy 556 ProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnThrProAla 575
 Db 1793 CCACATGTTAGTGCATTTGCTGTCTCAAGCCAGACATCCAGATTGGAGTCCAGCA 1852
 Qy 576 AlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIle 595
 Db 1853 GCAATTAATCCGCACTCATGACACCTGTTATGTTTCATGATAACACTTACAAATCTCTA 1912
 Qy 596 LysAspGlyAsnLys-----ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsn 613
 Db 1913 AAAGATCCCTCATCCGTGACTCTCTCGACCCCATATGATCACCCTGCAGGACACGTCAT 1972
 Qy 614 ProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThr 633
 Db 1973 CCAAGGAAGCGGTGTATCCGGTCTGATTATGACATAGGGGCACAAAGATTACTTCGAG 2032
 Qy 634 TyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIle---ThrHisLysAsn 652
 Db 2033 TTTCTGTGCACGCAAGAACTCAGCCCTTCACAGCTAATGGTTTTCGGCAAGTTTCTCAAC 2092
 Qy 653 ValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIle 672
 Db 2093 AGAATTCGCCACCTCAGCT---GCTAATCCAGG---GACTTGACTACCCAGGCATC 2146
 Qy 673 AlaValIlePheLysArgGlyLysThrThrGluMetIleThr-----ArgArgValThr 690
 Db 2147 TCTGAGTTTTCCTGAAAGACAAACTTTCAATGTGACCTTCATAGAACAGTCACT 2206
 Qy 691 AsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLys 710
 Db 2207 AATGTGGTTTCCCACTCAAAATTTACCATGTTGTAGTCTCAGCATTTTAAAGCGCAGTC 2266
 Qy 711 ValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArg 730
 Db 2267 GTGAAGTTGAGCAGACGATTCGAATTTCCACGACAAAACACAGAACTGTCTTCAAG 2326
 Qy 731 ValTyrPheValLeuLysLysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGly 750
 Db 2327 GTCATTTCAAAACCGTATCTCGTCAA-----AAAGCACTGAAATTT-----GGA 2371
 Qy 751 GlnLeuThrTyrValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal 770
 Db 2372 TCCCTAATATGGAAGATGGA-----ACACATAAGTAAGAGCCGCGGATGCAATC 2422
 Qy 771 Thr 771
 Db 2423 ACA 2425
 RESULT 12
 LRS6379
 LOCUS LRS6379 5197 bp DNA linear PLN 30-NOV-1999

DEFINITION Lycopersicon esculentum sbt2 gene.
 ACCESSION AJ006379
 VERSION 1 GI:3687306
 KEYWORDS sbt2 gene; serine protease; subtilisin-like protease.
 SOURCE Lycopersicon esculentum (tomato)
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 REFERENCE 1
 AUTHORS Meichtry, J., Amrhein, N. and Schaller, A.
 TITLE Characterization of the subtilase gene family in tomato (Lycopersicon esculentum Mill.)
 JOURNAL Plant Mol. Biol. 39 (4), 749-760 (1999)
 MEDLINE 9277592
 PUBMED 1035089
 REFERENCE 2 (bases 1 to 5197)
 AUTHORS Schaller, A.
 TITLE Direct Submission
 JOURNAL Submitted (26-MAY-1998) Schaller A., Institute of Plant Sciences, Federal Institute of Technology Zurich, Universitätsstr. 2, CH-8092 Zurich, SWITZERLAND

FEATURES
 source
 1..5197
 /organism="Lycopersicon esculentum"
 /mol_type="genomic DNA"
 /cultivar="VFw8"
 /db_xref="taxon:4081"
 1793..4446
 /gene="sbt2"
 /gene="sbt2"
 /gene="sbt2"
 /evidence="experimental"
 1908..4235
 /gene="sbt2"
 /function="serine protease"
 /codon_start=1
 /product="subtilisin-like protease"
 /protein_id="CAA07000.1"
 /db_xref="GI:3687307"
 /db_xref="SPRMBL:P93205"
 /translations="MAGMLLKMFVSVCLAINLAKSPNTKTVIIQMDKWKPDV
 FVDHVOYSLVKSVPSPTEVEKTDGGERILYSYQTAHPGVAQLSEEVKQLQER
 NGVAFFPEVLQHTFRPLFLGLDREDSKLDRLSDHNVIVGLDITWPSRPS
 FNDGMSVPEHWKVCETGPEFKHCKKIVGARVFRGVEAASGKINERGEFSA
 RDQDGHGTAGTVAGSVRGANLLGYCTARGAPVAAYKVVYWGCGFSSDIL
 SAVDQAVDQVNTLSISLGGVSSVNRDLSIAAFGAMKGVFVSCAGGQPDPIIL
 TNYSPMTTIVGASTMDRDPATVELCTGKIVTGSALYKGRNLSLTKQYPLIYLSNS
 SNLPSLCLDLGDKASVANGKIVCDRGISPRVQGVYKAGGVGMILNTANGE
 ELVADSHLUPAVAVGREGRAIKLYAAGRSATATLFLGLKIGIRPSPVVAFFSRGP
 NFLSLILKPDWVAGVNTLAGWTGALGPSLIPIDQRTNPNLSTGSMCPHVSIA
 ALLKARHPDWSPAIKSALMTAYVHDNTYKSLKASSVTPSTPYDHGAGHVNPRKAV
 DPLIYDIGAODYFEFLCTCELSPLSMVFGKFSNRTCHHSANPGDLNYPALSAVFP
 ETKLSMLTHRTVTVNGVPSISNHYVVSAPFGAVYKVEPERLNTSKNQLSKYVTF
 KTVSRQKAPFGLSLWKGTHKVRSPIALTWLASV"

prim_transcript <1793..4446
 /evidence="experimental"
 1908..4235
 /gene="sbt2"

CDS
 1908..4235
 /gene="sbt2"
 /function="serine protease"
 /codon_start=1
 /product="subtilisin-like protease"
 /protein_id="CAA07000.1"
 /db_xref="GI:3687307"
 /db_xref="SPRMBL:P93205"

BASE COUNT 1701 a 864 c 964 g 1668 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,37e-99 Length: 5197
 Score: 1679.00 Matches: 359
 Percent Similarity: 61.97% Conservative: 125
 Best Local Similarity: 45.97% Mismatches: 261
 Query Match: 41.79% Indels: 36
 DB: 8 Gaps: 16

US-09-806-767-2 (1-775) x LES6379 (1-5197)
 QY 6 PhePheLeuCysile-----IlePheLeuLeuPheCysSerSerSerSerGluile 22
 Db 1938 TTCTTTGATCAGTATCTCTGTATCACTTAGCCAAATGAGCCCAACACT----- 1991
 QY 23 LeuGlnLysGlnThrTrpIleValGlnLeuHisProAsnSerGluThrAlaLys----- 40

US-09-806-767-2 (1-775) x LES6379 (1-5197)
 QY 6 PhePheLeuCysile-----IlePheLeuLeuPheCysSerSerSerSerGluile 22
 Db 1938 TTCTTTGATCAGTATCTCTGTATCACTTAGCCAAATGAGCCCAACACT----- 1991
 QY 23 LeuGlnLysGlnThrTrpIleValGlnLeuHisProAsnSerGluThrAlaLys----- 40

Db 1992 -----AAGAAACTTACATTATCAAAATG-----GATAAATGGCTAAACACGAT 2036
 QY 41 ThrPheAlaSerLysPheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyVal 60
 Db 2037 GTATTGTCGACCATGTTCAATGTTACTCATCGTCAAAATCAGTACCTACCATCA 2096
 QY 61 GluGluGluGluGlu-----GluProSerSerArgLeuLeuTrpSerTrpGlySerAla 78
 Db 2097 ACAGAGGTAGAAAAACAGGGGATGGAGAGAGGATCTTTACAGCTACCACTGCT 2156
 QY 79 IleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluLeuLeuArgTrpSerPro 98
 Db 2157 TTTCATGCTGTGTCAGCTCAGCTTAGTGAAGACAGAGTGAATAAATTTACAAGACGAAT 2216
 QY 99 GluValValAlaValArgProAspHisValLeuGlnValGlnThrTrpSerTrpLys 118
 Db 2217 GGGGTTTTAGCTGTTTTCTCTGAGATTAAAGTATCAGTTATACCATCAGTAGAAGCTCTG 2276
 QY 119 PheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrpProGluSerProSerPheAsp 138
 Db 2277 TTTCTTGGCTTGACCGTAGGATAGCAGTAAGTTGTGGCGGATAGATTATCTGACCAT 2336
 QY 139 GlyThrIleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAsp 158
 Db 2337 AATGTAATTGCTGGTGTCTTGATCTGGGATTTGGCGGAAAGTCCGAGTTTCAATGAT 2396
 QY 159 ThrGlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPhe 178
 Db 2397 AGTGAATGACTAGTGTTCATCTCACTGGAAGGGGTTTGTGAACAGGAGGGGTTT 2456
 QY 179 SerSerSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArg 198
 Db 2457 GAGAAAGCATCATTCGACCAAGAAATTTGTTGGTCTAGAGTGTTCCTGGGTGATGAA 2516
 QY 199 ValAlaAsnSerProGluGluSerProAsnMetProArgGluTrpIleSerAlaArgAsp 218
 Db 2517 GCAGCTTCTGGT-----AAATCAATGAAGGGGTGAGTCAATCGGCAAGAGAT 2567
 QY 219 SerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAla 238
 Db 2568 CAAGATGGCATGCTACTCATCTGCTGGAACAGTTCGAGGTCTGTAGTCCGTGGTGT 2627
 QY 239 AsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAla 258
 Db 2628 AATCTTTGGGTATGCTTATGTTGCTAGTACTGCTAGAGAAATGCCCTGTGGCAGAGTTGCA 2687
 QY 259 ValTrpLysValCysTrpPheAsnGlyCysTrpSerSerAspIleLeuAlaAlaLeuAsp 278
 Db 2688 GCATATAGGTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2747
 QY 279 ValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIlePro 298
 Db 2748 CAAGCTGGCTGATGGGTAAATATCTGTCAATTTCTTGGGTGGTGGGTGGTGGTGGT 2807
 QY 299 LeuTrpAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerVal 318
 Db 2808 TATAATCGTGACAGTTTCTCAATTCGCTTTTGGGGCCATCGAGAAAGGGGTTTTGTT 2867
 QY 319 IleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrp 338
 Db 2868 TCATGTTTCAGCAGAAATGGTGGCCCTGATCCCAATTAGTTGACAAATGTGTCCACATGG 2927
 QY 339 ValSerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAla 358
 Db 2928 ATCACTACTGTAGAGCAAGCACTATGATAGAGATTTTCCAGCAACTGTTGACCTGGGG 2987
 QY 359 AsnGlyLysLeuLeuTrpGlyGluSerLeuTrpProGlyLysGlyIleLysAsnAlaGly 378
 Db 2988 ACGGGAGAGATTGACCGGAGCTTCACTTTATAAGAGTAGGATGAATCTTCAACACAA 3047
 QY 379 ArgGluValGluValIleTrpValThrGlyGlyAspLysGly-----SerGluPhe 395

Db 3048 AAACAGTATCCCTTTGATTACTTAGGAAGTAATTCGAGCAATCTCATGCCGAGCTCAATTG 3107
 Qy 396 CyeLeuArgGlySerLeuProArgGluGluIleAArgGlyVlyMetValIleCysAspArg 415
 Db 3108 TGCTTAGATGGTACCTTAGATAAGCTTCAGTTGCTGGAAATAATGTGATATGTGCACAGG 3167
 Qy 416 GlyValAsnGlyArgSerGluGlyGlyGluAlaValIleGlyAlaGlyGlyValAlaMet 435
 Db 3168 GGAATAGTCTCGAATTCMAAGGGGCAAGTTGTAAAGAAGCTGGAGGAGTAGAATG 3227
 Qy 436 IleLeuAlaAsnThrGluIleAsnGlnGluLeuAspSerIleAspValHisLeuLeuPro 455
 Db 3228 ATTTTGACAAACACTGCAGCTAATGTGAAGAACTGGTTGCTGATTCTCACCTTCTTCG 3287
 Qy 456 AlaThrLeuIleGlyTyrThrGluSerValLeuLeuLeuLeuAlaTyrValAsnAlaThrVal 475
 Db 3288 GCAGTAGCAGTGTGTGAAGAGAGGAAGAGCAATCAAGCTCTACGCG---GCAGGTAGA 3344
 Qy 476 LysProIleAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
 Db 3345 AGCGTACTCCACTTTGAGATTCTTGGAATTTCTGGACTAAATTAGGATCAGGCTTCGCCGTG 3404
 Qy 496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515
 Db 3405 GTTGTCTGCATTTTCATCCAGAGGACCTAATCTCTCAGCTAGAGATTCTAAGCCAGAT 3464
 Qy 516 MetIleAlaProGlyValAsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGly 535
 Db 3465 ATGGTGCCACGAGGTGAATATCTTGTGCTGTTGCTGATGCTGCTGCTGCTGCTGCTGCT 3524
 Qy 536 LeuProTyrAspSerArgValAsnPheThrValMetSerGlyThrSerMetSerCys 555
 Db 3525 TTCCCAATTGATCAACGAGGACGAATTTCAATATTCTCTGGAACCTTCAATGTCATGT 3584
 Qy 556 ProHisValSerGlyIleThrAlaLeuIleAArgSerAlaTyrProAsnThrProAla 575
 Db 3585 CCACATGTTAGTGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3644
 Qy 576 AlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIle 595
 Db 3645 GCAATTAATCCGCACTCATGACCACTGCTTATGTTATGATGATGATGATGATGATGATGAT 3704
 Qy 596 LysAspGlyAsnLys-----ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsn 613
 Db 3705 AAAGATGCCCTCCGCTGACTCTCTGACCCCACTATGATCAGCGTCAGGACAGCTCAAT 3764
 Qy 614 ProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThr 633
 Db 3765 CCAAGGAAGCGGTGATCCCGTCTGATTTATGATGATGATGATGATGATGATGATGATGATGAT 3824
 Qy 634 TyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIle---ThrHisLysAsn 652
 Db 3825 TTCTTGTGACGCAAGCAAGCACTCAGCCCTTCACAGCTAATGTTTTCGGCAAGTTCTCAAC 3884
 Qy 653 ValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIle 672
 Db 3885 AGAATTCGCCACCACTCACTT---GCTAATCCAGGG---GACTTGAACCTACCCAGCCATC 3938
 Qy 673 AlaValIlePheLysArgGlyLysThrThrGluMetIleThr-----ArgArgValThr 690
 Db 3939 TCTGCAGTTTCTCTGAAAAGACAAACTTTCATGCTGACCTTCATGAAACAGTCACT 3998
 Qy 691 AsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLys 710
 Db 3999 AATGTGGTTTCCCCCATCTCAATATCATGTTGTAGTCTCAGCATTTAAAGCCGAGCTC 4058
 Qy 711 ValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArg 730
 Db 4059 GTGAAGGTTGAGCCAGAACGATTGAAATTTCCACCAGCAAAACCCAGAACTGCTTTTACAAG 4118
 Qy 731 ValTrpValLeuLeuLysLysLysAsnArgGlyGlyValAlaLysPheAlaGlnGly 750
 Db 4119 GTCACATTCAAAGCGGTATCTCGTCAA-----AAAGCACCTGAATTT-----GGA 4163

Qy 751 GlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal 770
 Db 4164 TCCCTAATATCGAAGATGGA-----ACACATAAGTAAGAGCCGATTCGAATC 4214
 Qy 771 Thr 771
 Db 4215 ACA 4217
 RESULT 13
 AY142612 2305 bp mRNA linear PLN 23-SEP-2002
 LOCUS Arabidopsis thaliana putative subtilisin serine protease ARA12
 DEFINITION (At5g67360) mRNA, complete cds.
 ACCESSION AY142612 GI:23296831
 VERSION AY142612.1
 KEYWORDS FUJ CDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 2305)
 AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W.,
 Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
 Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R.,
 Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J.,
 Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
 Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K.,
 Davis,R.W., Ecker,J.R. and Theologis,A.
 TITLE Arabidopsis Open Reading Frame (ORF) Clones
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2305)
 AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W.,
 Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K.,
 Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R.,
 Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J.,
 Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
 Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K.,
 Davis,R.W., Ecker,J.R. and Theologis,A.
 TITLE Direct Submission
 JOURNAL Submitted (16-AUG-2002) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 COMMENT The RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
 Arabidopsis Full-length cDNA': Seki,M., Narusaka,M., Ishida,J.,
 Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
 Hayashizaki,Y. and Shinozaki,K.
 The Salk, Stanford, PDEC (SSP) Consortium members constructed and
 sequenced the PUNI (ORF) clones using the RAFL cDNAs: Yamada,K.,
 Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hsuan,V.W., Lee,J.M.,
 Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
 Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Kim,C.J.,
 Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T.,
 Davis,R.W., Ecker,J.R. and Theologis,A.
 Yamada,K. (SSP/PDEC) and Seki, M. (RIKEN GSC) contributed equally
 to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP
 /PDEC) contributed equally to this work as PIs.
 Annotation is based on the January 2002 version of the Arabidopsis
 genome submitted to GenBank.
 FEATURES
 source location/Qualifiers
 1..2305
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="U09277"
 /note="This clone is in pUNI 51.
 ecotype: Columbia"
 1..2305
 gene

```

/ gene="At5g67360"
1. .2274
/ gene="At5g67360"
/ note="contains subtilase domains PF00082 and a PA domain
PF02225"
/ codon_start=1
/ evidence=experimental
/ product="putative subtilisin serine protease ARA12"
/ protein_id="AA13181.1"
/ db_xref="GI:23296832"
/ translations="MSSSLSTAFPLLLCLGFCVSSSSDQGVIVHMAKSWMPSS
FDLHSWYDSLSRISDSAEALLTYENAHGFSRLTQBEADSLMTQPVISLPEHR
YELHTPLFLGDSHEDLDFEAGSYSDVVVLDLTQWPEKSYSDGFGPIPS
WKGCSAGTINFLASLCKRLIGARFFARGVETMGPIDESKSRPRDDHGTST
TAAGSVGAGSLGASGTARGMAPRVAIVKVLGGCFSDILADKAIADNVN
VLSMLGGMSDYVRDQVGAFAAMERGLVSCSAGNAGPSSSLSNVAPWITVGA
GLTDROPALAILGNQKPTGSLPKGEALPKLLPIYAGNASNATGNLGMTGLI
PEVKKIKVMDGRNARVQGVVKAAGVGMILANTANGELVADHLLPALTIVG
EKAGDIRHVVTDPNPTASISILGTWVGKSPVMAFFSPNSITNHLKPLIA
PGVILAAATGAAGPTGLASDSRVFNIIISGTMSCPHVSLAALLKSVHPWSPAA
IRGALMTAYTKVDKGLDIDATGKPSYDFOHGAGVSPVATNPLGLYDLTLDYL
IFCALNYSPOIRSVRNRYTCDPSKSYVDLNLNPSFAVNDVGVAKYTRTVSV
GGAGTYSVKTSITGKISVEPAVLNFKENAKKSYTVTFVDSKPSGNSPSGIE
WSDGKHVVGSPVAISWT"
2275..2305
/ gene="At5g67360"
BASE COUNT 547 a 567 c 580 g 611 t
ORIGIN

Alignment Scores:
Pred. No.: 2,93e-99 Length: 2305
Score: 1666.50 Matches: 362
Percent Similarity: 61.97% Conservative: 122
Best Local Similarity: 46.35% Mismatches: 246
Query Match: 41.48% Indels: 51
DB: 8 Gaps: 18

US-09-806-767-2 (1-775) x AV142612 (1-2305)
QY 6 PhePheLeuCyslellePheLeuPheCysSerSerSerSerGluLeuLeuLys 25
DB 31 TTCTTCCTC---CTCCTTTCTAGGATTCGCCAGCTGTCTCTCTCTCTCCCGACCAA 87
QY 26 GlnThrTyrlleValGlnLeuHisProAsnSerGluThrAlaLysThr---PheAlaSer 44
DB 88 GGAACCTTACATCGTTCACATG-----GCTAAATCTCAGATGCCGCTG 129
QY 45 LysPheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGlu 64
DB 130 TCTTTCGACCTCCACTACTAGTACGATTCATCTCTCAGATCTATCTCCGAC----- 183
QY 65 GluGluProSerSerArgLeuLeuTyrlleValSerTyrlleValSerAlaAlaAla 84
DB 184 -----TCAGCGCAATTGCTCTACACTTACGAGACGCGATTCATGATCTCAACA 234
QY 85 GlnLeuThrGluSerGluAlaGluLeuLeuArgTyrlleValValAlaValArg 104
DB 235 CGGCTAACTCAAGAGAGCGCACTCGCTCATGACTCAACCTGGTGTATCTCCGTTT 294
QY 105 ProAspHisValLeuGlnValGlnThrThrTyrlleValSerTyrlleValSerGly 124
DB 295 CCGGACACCGGTACGAGCTACACACACTCGTACTCTCTCTCTCTCTCTCTCTCT 354
QY 125 -----PheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGly 139
DB 355 CACACCGCAGATCTGTTCCCTGGAAGCGCGCTCTTACAGC-----GAC 396
QY 140 ThrIlelleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspThr 159
DB 397 GTGCTGCTCGGAGTCTTCGATACCGGAGTTGCGCGGAGTAAAGAGTACTCCGACGA 456
QY 160 GlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGluGluSerPheSer 179
DB 179 ATCGCTCTGAGTAAACATCTCTCGCGCGGTGGACCGGTCTGCTGGACCAACCGGACTC 1578

```

```
QY 537 ProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysPro 556
Db 1579 GCTTCGGATCTCGCGCGGTGGAGTTCAATATCTCTCGGACGTCGATGTTGGCCCT 1638
QY 557 HisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTyrSerProAlaAla 576
Db 1639 CAGGTAGTGGTTAGCGCGGTTCTCAAGTCGGTGCATCTCGATCGAGCCGCGCGG 1698
QY 577 IleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyAlaIleLys 596
Db 1699 ATTAGATCGCGGTTATGACCGGTTACAAACCTACAAAGACGGTAACCGTTACTC 1758
QY 597 Asp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro 614
Db 1759 GACATCGGACGAGGGAAGCTTCGACCGGTTTCGATCAGCGTGCAGGACACGTTGCACCA 1818
QY 615 GlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyr 634
Db 1819 ACMACTGCCACTATCCAGACTCATCTACGATCTAACGACGAGATTACTAGGCTTC 1878
QY 635 LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSer 654
Db 1879 CTCTGCGCATTAATACATCGCGCAAAATTCGAAGTGTCTCGAGACGTAATTACACT 1938
QY 655 CysAsnGlyIleLeuArgLysAsnProGlyPheSer-----LeuAsnTyrProSer 671
Db 1939 TGGCAT-----CCGAGTAAATCGTACTCGTGGTGGTGAATTTGAACCTACCCGTCG 1986
QY 672 IleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArgArgValThrAsn 691
Db 1987 TTGCGCGTTAACTGGAT---GGAGTCGGTGGGTATAGTACACGCGACGCGTACGAGC 2043
QY 692 ValGlySerProAsnSerIleTyrSerValAsnValLysAla---ProGluGlyIleLys 710
Db 2044 GTGGGAGGAGCTGGGACT---TACTCGGTTAAGTAACCTCGGACGACGAGTCAAG 2100
QY 711 ValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArg 730
Db 2101 ATTTCGGTTGAACCGCGGTTTGAATTTCAAGGAGCTAACGAGAAAAATCGTATACG 2160
QY 731 ValTyrPheValLeuLysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGly 750
Db 2161 GTGACGTTTACTGTAGACTCGTCAAGCCGCTCTGGA-----TCTAACGAGCTTGGG 2211
QY 751 GlnLeuThrTyrValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal 770
Db 2212 AGTATTGAATGGTCGGATGGGAAACACGTTG-----GTGGGAATGCCCGTGGGATT 2262
QY 771 Thr 771
Db 2263 AGC 2265

RESULT 14
AF360285
LOCUS
DEFINITION
Arabidopsis thaliana putative subtilisin serine protease ARA12
(AT5g67360) mRNA, complete cds.
ACCESSION
AF360285
VERSION
AF360285.1 GI:13430745
KEYWORDS
FLI CDNA.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 2618)
REFERENCE
Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
2389. .2618
/gene="At5g67360"

3'UTR
```

Davis,R.W., Ecker,J.R. and Theologis,A.
Arabidopsis Full Length cDNA Clones

TITLE
JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 2618)
Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
Davis,R.W., Ecker,J.R. and Theologis,A.
Direct Submission
Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

TITLE
JOURNAL

COMMENT

The Salk, Stanford, PGECC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,
Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D.,
Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Chen,H.,
Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A.,
Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGECC) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGECC)
contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to Genbank.

FEATURES
source

1..2618
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAFL05-16-L13 (R09277)"
/note="This clone is in a modified pBluescript vector
(pUC-1) as a BamHI/XhoI insert.
ecotype: Columbia"

gene
5'UTR
CDS

1..2618
/gene="At5g67360"
1..114
/gene="At5g67360"
115..2388
/gene="At5g67360"
/note="contains subtilase domains PF00082 and a PA domain
PF02225"
/codon_start=1
/evidence=experimental
/product="putative subtilisin serine protease ARA12"
/protein_id="AAK25995.1"
/db_xref="GI:13430746"
/translation="MSSSFLSSTAFFLLCLGFCVSSSSDQGTIVHMAKSMFSS
FDLHNWYDSSLSRISDSABELLYTENAIHGFSRTRTOEADSLMTQPGVIFLPEHR
YELHTRTPLEGLDEHATDFEAGSYSDVVGVDLTGVNPEKSYSDGSGFPSS
WKGCEAGTFTASLCNRKLGARFARFARGVSTMGPIDESKSRSPRDDDGHGTHRS
TAAGSVVEGASLIGYAGTARGMAFPAARVAVYKVGWGGCFSSDILAAIDKALADNVN
VLSMGLGNGSYRDGVAIGAFPAAMERGLIVCSAGNAPSSSSNSVNPWITTVGA
GTLDKDPALAILNGNFTGSLFKGELPKLLPFIYAGNASNATNGMLCMTGLI
PEVKGKIIVMCDRGINARVCKGVKAAAGVGVMILANTANGELVADAHLLPATTVG
EKAGDIIRHVTTDPTASISILGTGVVGVKSPVVAAPSRRGPNSTPNILKPDLIA
PGVNLIAAATGAAGPTGLASDSRRVFNIISGSPSCMSPHVSGLAALLKSVHPSPNA
IRGALMTAYTKYDKGKPLDIDATGKSPFPFDHGAGHVSPTTATNGLIYDLTETDL
GFLCALMTYQPSRSRRNYTCDPSKSYSDADLNTPSPAVNVDGAGKYKTTVTSV
GGAGTSYKVTSETTGTGKISVEPAVLNFKANEKSYVTFTTVDSSKPSGSGNSFGSIE
WSGKHVVGSPVAISWT"
2389. .2618
/gene="At5g67360"

```

misc_difference 2602
/genes="AT5g67360"
/notes="not present in genomic sequence"
BASE COUNT      642 a   632 c   628 g   716 t
ORIGIN

Alignment Scores:
Pred. No.:      3,49e-99      Length:      2618
Score:          1666.50      Matches:      362
Percent Similarity: 61.97%      Conservative: 122
Best Local Similarity: 46.35%      Mismatches:   246
Query Match:      41.48%      Indels:       51
DB:                8          Gaps:         18

US-09-806-767-2 (1-775) x AF360285 (1-2618)

Qy      6 PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLys 25
Db      145 TTCTTCCTC---CTCCTTTGCTAGGATTCTCCACAGTGTCTCTCTCTCTCTCTCCGACCAA 201
Qy      26 GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThr---PheAlaSer 44
Db      202 GGAACCTTACATCGTTCCATG-----GCTAAATCTCAGATGCCGTCG 243
Qy      45 LysPheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGlu 64
Db      244 TCCTTCGACCTCCACCTTAACCTGGTACGATTCTCTCAGATCTATCTCCGAC----- 297
Qy      65 GluGluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAla 84
Db      298 -----TCAGCCGAATTGCTTACACTTACGAGACGCGATTCTGATTTCTCAACA 348
Qy      85 GlnLeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArg 104
Db      349 CGGCTAACTCAAGAAAGCCGACTCGCTCATGATCAACCTGGTGTATCTCCGTTTAA 408
Qy      105 ProAspHisValLeuGlnValGlnThrTyrSerTyrLysPheLeuGlyLeuAspGly 124
Db      409 CCGAGACACGGTTACAGGTACACACCTCGTACTCTCTCTCTCTCTCTCTCTCTCTCTCT 468
Qy      125 -----PheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGly 139
Db      469 CACACCCGACAGATCTGTTCTCCGAAAGCGGCTCTTACAGC-----GAC 510
Qy      140 ThrIleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspAspThr 159
Db      511 GTCTGCTCGGAGTTCTCATACCGGAGTTTGGCCGAGATAAAGACTACTCCGACGAA 570
Qy      160 GlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyCyluSerPheSer 179
Db      571 GGATTCGGTCTTATCTCTCTCTGAAAGCGGATGCGAGCGCGAACCAATTTTACC 630
Qy      180 SerSerSerCysAsnArgLysLeuIleGlyValAlaArgPhePheIleArgGlyHisArgVal 199
Db      631 GCTTCTCTCTGTACCGTAACTAATCGAGACAAGATTTCTCGCTCGGTGTACGAATCA 690
Qy      200 AlaAsnSerPro---GluGluSerProAsnMetProArgGluTyrIleSerAlaArgAsp 218
Db      691 ACGATGGGACCAATCGATGAATCT-----AAAGATCAAGATCTCTCTAGAGAC 738
Qy      219 SerThrGlyHisGlyThrHisThrAlaSerThrValGlyLysSerSerValSerMetAla 238
Db      739 GAGCAGCGGTCAACGAACTCACACCTCATCGCGCGCTGGATCGCTGTGAAGAGCT 798
Qy      239 AsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyValAlaHisIleAla 258
Db      799 AGCTTATAGGTACCGCTTCAGGAACAGCTCTGTGGTATGGCTCCACGCGCTCGTCTGCT 858
Qy      259 ValTyrLysValCysTrpPheAsnGlyCysTrpSerSerAspIleLeuAlaIleAsp 278
Db      859 GTTTACAAAGTCTGTGGTGGTGGTGTGTTTCAGCTCAGATATTTAGCTGCATTCGAT 918
Qy      279 ValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyPheProIlePro 298

```

```

Db      919 AAAGCCATCGCCGATACGTCAATGCTGTGTCATGTCACCTCGCGTGGAAATGTCGGAT 978
Qy      299 LeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerVal 318
Db      979 TATTATAGACACGGTGTTCGATCGGAGCAATTCGCCCATGGAAGAGGAGATTTAGTA 1038
Qy      319 IleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrp 338
Db      1039 TCTTGCTCAGCTGGTAAATGCTGTCGAGTAGTCTTAGTTTATCAAAAGTAGTCCATGG 1098
Qy      339 ValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAla 358
Db      1099 ATCACAACCTGCTGCTGCTGCTACTTAGATCGTGATTTTCCGCGCTTCCGATTTCTCGGT 1158
Qy      359 AsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGly 378
Db      1159 AACGGCAAGAATTTCAACCGAGTTCTTTGTTTAAAGGAGAGCTTTTACCGGAT----- 1212
Qy      379 ArgGluValGluValIleTyrValThrGlyLysGlySer-----GluPheCys 396
Db      1213 AAATGTTGCGGTTTATTACGCTGGGAATGCTAGTAATGCTACTAATGGTATCTCTGT 1272
Qy      397 LeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArgGly 416
Db      1273 ATGACCGCAACTTTGATCCCGGAGAAAGTAAGGGAAGATTGTGATGTGTATAGAGA 1332
Qy      417 ValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyIleValAlaMetIle 436
Db      1333 ATTAATCTAGATTTCAGAAAGGTGATGTGTTTAAAGCAGCAGTGGAGTTGGAATGATT 1392
Qy      437 LeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeuLeuProAla 456
Db      1393 CTGGCTTAACCTCGCGGAATGTTGAAGACTTGTTCGCGATGCTCATTTGTTACCGGCG 1452
Qy      457 ThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLys 476
Db      1453 ACCACCGTTGGTGAAGAAAGCGGTATATAATCCGCAATACCTGACTACTGATCTAAT 1512
Qy      477 ProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGluVal 496
Db      1513 CCCACCGCTTCGATTTCAATCTTAGGAACAGTCTGGTGTAAACCATCTCCGGTTGTC 1572
Qy      497 AlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMet 516
Db      1573 GCAGCGTTTACGTACGTGACGTCGAATTCGATTACACCAATATCTTAAACCGGATCTG 1632
Qy      517 IleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeu 536
Db      1633 ATCGTCTCGAGTAAACATCTTCGCGGGTGGACCGGTGCTGCTGGACCAACCGGACTC 1692
Qy      537 ProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysPro 556
Db      1693 GCITTCGATTTCTCGCGGTGGAGTTCAATATCATCTCTGGGACGTCGATGTTCTGCCCT 1752
Qy      557 HisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAla 576
Db      1753 CAGGTAGTGGTTAGCGGCGCTTCTCAAGTCGTCATCTCTCAATGAGCGCGCGGCGG 1812
Qy      577 IleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLys 596
Db      1813 ATTAGTCGCGGCTTATGACCCCGTTTCAAAACCTTACAAACGCGGTAAACCGTTACTC 1872
Qy      597 Asp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro 614
Db      1873 GACATCGCGACAGGAGCGCTTCGACGCGGTTGATCATCGGTGCAGGACACGTTGTCACCA 1932
Qy      615 GlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyr 634
Db      1933 ACRACTCCCACTAATCCAGGACTCATCTAGATCTAACGACGAGGATTACTTAGGCTTC 1992
Qy      635 LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSer 654

```


Qy 180 SerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgVal 199
GCTTCTCTGTAAACCGTAATAACTAATCGGAGCAAGATTCTTCGCTCGTGTTACGAATCA 865

Dy 200 AlaAsnSerPro---GlUGluSerProAsnMetProArgGluTyrlleSerAlaArgAsp 218
ACGATGGAGCCAAATCGATGAATCT-----AAAGAATCAAGATCTCTTAGAGAC 913

Qy 219 SerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAla 238
GACGACGGTCACGGAACTCACACCTCATCGACCGCGCTGGATCCGTCGTGAAGAGACT 973

Qy 239 AsnValLeuGlyAsnGlyAlaGlyValAlaAArgGlyMetLaProGlyAlaHisIleAla 258
AGCTTATTAGCTACGGTTCAGGAACAAGCTCGTGGTATGGCTCCACGCGCTCGTGCCT 1033

Qy 259 ValTyrlleValCysTrpPheAsnGlyCysTrpSerSerAspIleLeuAlaIleAsp 278
GTTTTACAAAGTCTGTGGCTCGGTGGTGTTCAGCTCAGATATTTAGTGTGCAATCGAT 1093

Qy 279 ValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIlePro 298
AAAGCCATCGCCGATAACGTCATGTCTCTCGATGTCACTCGGCGGTGGATTCGGAT 1153

Qy 299 LeuTyrlleAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerVal 318
TATTATAGACACGGTGTTCGATCGGAGCATTCGCCCATCGGAAGAGGATTTAGTA 1213

Qy 319 IleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrp 338
TCTTGTCTCAGCTGTGTAATGCTGGTCCGAGTAGTTCATTAGTTCATAAACGCTAGTCCATGG 1273

Qy 339 ValSerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValAlaValGLeuAla 358
ATCACAACTGTGTGTCTGTGTACTTTAGATCGTGAATTTCCGGCGCTTCGGATTCCTGGT 1333

Qy 359 AsnGlyLysLeuLeuTyrlleGlyGluSerLeuTyrlleGlyLysGlyIleLysAsnAlaGly 378
AACGGCAAGAAATTTACCGGAGTTCTTTGTTTAAAGGAGAGCTTTTACCGGAT----- 1387

Qy 379 ArgGluValGluValIleTyrlleValThrGlyGlyAspLysGlySer-----GluPheCys 396
AAATGTGTTCGGTTTATTACGCTGGAAATGCTAGTAATGCTACTAATGTAATCTCTGT 1447

Qy 397 LeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArgGly 416
ATGACCGGAACCTTGTATCCCGAGAAAGTAAAGCGGAAGATGTGATGTGTGATAGAGGA 1507

Qy 417 ValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIle 436
ATTAAATGCTAGAGTTTCAGAAAGGTGATGTGGTTAAAGCAGCAGGTGGAGTTGGAATGATT 1567

Qy 437 LeuAlaAsnThrGluIleAsnGlnGluLysAspSerIleAspValHisLeuLeuProAla 456
CTGGCTTAACACTCGCGCAAATGTGTGAAGAGCTTGTTCGGATGCTCATTTGTTACC GGCG 1527

Qy 457 ThrLeuIleGlyTyrlleThrGluSerValLeuLeuLysAlaTyrlleValAsnAlaThrValLys 476
ACACCGTGTGTGAAAAAGCCGGTGATATTAATCCGGCATTAACGTGACTACTGATCTTAAT 1687

Qy 477 ProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGluVal 496
CCCACCGCTTCGATTTCAAATCTTAGGAACAAGTCGTGGTGTAAACCATCTCCGGTTGTC 1747

Qy 497 AlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMet 516
GCACGCTTATAGCTCACGTGACCGCAATTCGATTACACCGAATATCTTTAAACCGGATCTG 1807

Qy 517 IleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeu 536
ATCGCTCTCGAGTPAALAAATCTTCGCGCGGTGACCGGTGCTGCTGACCAACCGGACTC 1867

Qy 537 ProTyrlleAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysPro 556

Db	1868	GCTTCCGATTCTCGCGCGTGGAGTTCAATATCATCTCTGGCAGCTGCATGTCTTGCCT	1927
Qy	557	HisValSerGlyIleThrAlaIleuIleArgSerAlaTyrProAsnTrpSerProAlaA	576
Db	1928	CACGTTAGTGGTTTAGCGGCGCTCTCAAGTCGGTCGATCCTGAATGGAGCCGGCGGG	1987
Qy	577	IleLysSerAlaIleuMetThrThrAlaAspLeuTyrAspArgGlnGlyAlaIleLys	596
Db	1988	ATTAGATCGCGGCTTATGACCACCGTTACAAACCTACAAAGACGGTAACCGTTACTC	2047
Qy	597	Asp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro	614
Db	2048	GACATCGCAGACGGAGACCTTCGACCGCTTCGATCAGGTGCAGACACAGGTGCACCA	2107
Qy	615	GlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyr	634
Db	2108	ACAACTCCCACTAATCCAGAGACTCATTCAGATCTAACGACGGAGGATTACTTAGCTTC	2167
Qy	635	LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSer	654
Db	2168	CTCTGCCCATTTGAATTACACATCGCGCGAAATTCGAAGTGTCTCGAGACGTAATTACAT	2227
Qy	655	CysAsnGlyIleLeuArgLysAsnProGlyPheSer-----LeuAsnTyrProSer	671
Db	2228	TGGCAT-----CCGAGTAATCGTACTCGTCTCGTCAATTGAACTACCGCTCG	2275
Qy	672	IleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArgArgValThrAsn	691
Db	2276	TTGCGCGTTAACGTTGAT--GGAGTCGGTGGCTATAGTACACGGCGCAGGTGACGAGC	2332
Qy	692	ValGlySerProAsnSerIleTyrSerValAsnValLysAla---ProGluGlyIleLys	710
Db	2333	GTGGAGGAGCTGGGACT--TACTCGGTTAAAGTAACCTTCGGAGACGACAGAGTCAAG	2389
Qy	711	ValIleValAsnProLysArgGluValPheLysHisValAspGlnThrLeuSerTyrArg	730
Db	2390	ATTTCGGTTGAACCGCGGGTTTGAATTTCAAGGAAGCTAACAGAGAAAATCGTATACG	2449
Qy	731	ValTrpPheValLeuLysLysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGly	750
Db	2450	GTGACGTTTACTGTAGACTCGTGAACCGCTCTGGA-----TCTAACACGCTTGGG	2500
Qy	751	GlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal	770
Db	2501	AGTATTCAATGGTGGATGGGAAACACGTG-----GTGGGAAGTCCCGTGGCATT	2551
Qy	771	Thr	771
Db	2552	AGC	2554

Search completed: February 3, 2004, 03:56:37
Job time : 5471 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 19:57:04 ; Search time 445 Seconds
(without alignments)
14090.340 Million cell updates/sec

Title: US-09-806-767-1

Perfect score: 2328

Sequence: 1 atggaacccaacaccttctt.....taacctgaagactaactga 2328

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2328	100.0	2328	21	AAZ93808 SDD1 sequence enco
2	2326.4	99.9	2328	21	AAZ93809 SDD1 sequence enco
3	1024.6	44.0	2492	21	AAZ93813 Subtilase (Pigesp)
4	1020.2	43.8	3140	21	AAZ93814 Subtilase (St_P2Sc
5	602	25.9	3865	21	AAZ93812 Subtilase (Pigesp)
6	481.8	20.7	2265	21	AAZ93317 Arabidopsis thalia
7	481.8	20.7	2265	24	AB213260 Arabidopsis thalia
8	447.4	19.2	2434	21	AAZ47609 Arabidopsis thalia

9	424.4	18.2	2637	21	AAZ46158 Arabidopsis thalia
10	417	17.9	2638	21	AAZ42012 Arabidopsis thalia
11	398.2	17.1	2295	24	ABZ13381 Arabidopsis thalia
12	398.2	17.1	2489	21	AAZ42230 Arabidopsis thalia
13	285	12.2	1653	21	AAZ32938 Arabidopsis thalia
14	228.2	9.8	2402	21	AAZ42099 Arabidopsis thalia
15	174.6	7.5	740	24	ABQ65482 Arabidopsis thalia
16	168.2	7.2	2552	15	AAQ73756 Gene encoding melo
17	123.2	5.3	436	25	ABX61910 Arabidopsis thalia
18	121.2	5.2	2310	25	ABZ42107 Arabidopsis thalia
19	113.6	4.9	2448	20	AAZ29905 cDNA encoding a SC
20	110.4	4.7	975	21	AAZ49798 Arabidopsis thalia
21	100.2	4.3	2958	24	ABZ12533 Arabidopsis thalia
22	90.8	3.9	513445	22	AAI61373 Soybean 318013 reg
23	83.4	3.6	363	25	ABX20410 Human GDP-mannose
24	82	3.5	7235	20	AAZ29910 Genomic DNA encodi
25	78.8	3.4	343	25	ABX21785 Human GDP-mannose
26	71.2	3.1	279	25	ABX23353 Human GDP-mannose
27	67.8	2.9	401	25	ABX23353 Human GDP-mannose
28	65	2.8	513445	22	AAI61373 Soybean 318013 reg
29	63	2.7	528	25	ABX57523 Arabidopsis thalia
30	62.4	2.7	297	24	ABL76753 Corn tassal-derive
31	62.4	2.7	379	25	ABX20319 Human GDP-mannose
32	58	2.5	293	25	ABX88421 Corn ear-derived p
33	55.8	2.4	2451	24	ABZ12259 Arabidopsis thalia
34	51	2.2	255	25	ABX24387 Human GDP-mannose
35	49	2.1	273	24	ABL73078 Human GDP-mannose
36	44.8	1.9	1236	20	AAZ05920 Hyperthermostable
37	44.8	1.9	1566	18	AAZ85668 Pyrococcus furiosu
38	44.8	1.9	1962	18	AAZ85695 Hyperthermostable
39	44.8	1.9	1962	20	AAZ05929 Pyrococcus furiosu
40	44.4	1.9	295	25	ABX88947 Corn ear-derived p
41	43.6	1.9	454	24	ABL33821 Arabidopsis thalia
42	42.6	1.8	4590	22	AAZ24065 Yeast ADP9604-asso
43	41.6	1.8	531	25	ABX57480 Arabidopsis thalia
44	41.2	1.8	272	24	ABL72301 Corn tassal-derive
45	40.2	1.7	569	25	ABX56955 Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAZ93808
ID AAZ93808 standard; DNA; 2328 BP.
XX
AC AAZ93808;
XX
DT 16-AUG-2000 (first entry)
XX
DE SDD1 sequence encoding a subtilisin like serine protease.
XX
KW SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomach; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;
KW crop protection; feed; foodstuffs; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
CDS 1..2328
FT /*tag= a
FT /product= SDD1 subtilisin-like serine protease.
XX
PN WO200022144-A2.
XX
PD 20-APR-2000.
XX
PF 12-OCT-1999; 99WO-EP07633.
XX
PR 12-OCT-1998; 98EP-0119244.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

Db 601 AATTCACGAGGAATCAACAAACATGCTCTGTGAATACATTTCCGCAAGAGATTCAACG 660
Qy 661 GGACACGGGACTCACACCGCTCAACAGTTGGTGGATCCTCTGTTTCGATGCGGAATGTT 720
Db 661 GGACACGGGACTCACACCGCTCAACAGTTGGTGGATCCTCTGTTTCGATGCGGAATGTT 720
Qy 721 CTTGGCAATGAGAGCTGGTGGTCTGTGGATGGCTCTGGAGTCAACATTCAGTCTAT 780
Db 721 CTTGGCAATGAGAGCTGGTGGTCTGTGGATGGCTCTGGAGTCAACATTCAGTCTAT 780
Qy 781 AAGTCTGTGGTCAATGGTGTGTACAGTCTGACATCTAGCAGCTATAGATGAGG 840
Db 781 AAGTCTGTGGTCAATGGTGTGTACAGTCTGACATCTAGCAGCTATAGATGAGG 840
Qy 841 ATTCAAGATAAAGTCAATGGTGTGTACAGTCTGACATCTAGCAGCTATAGATGAGG 900
Db 841 ATTCAAGATAAAGTCAATGGTGTGTACAGTCTGACATCTAGCAGCTATAGATGAGG 900
Qy 901 GATGACACATCCCATTTGGAACATTCGAGCCATGGAACGCGGTATATCTGTAATCTGT 960
Db 901 GATGACACATCCCATTTGGAACATTCGAGCCATGGAACGCGGTATATCTGTAATCTGT 960
Qy 961 GCAGCTGTGTAACAAACGGTCCATCGAAAGCTCTGTTGCAACACAGCTCTTGGTCTCA 1020
Db 961 GCAGCTGTGTAACAAACGGTCCATCGAAAGCTCTGTTGCAACACAGCTCTTGGTCTCA 1020
Qy 1021 ACCATTGGCGCAGCAGCTGTGATCGAAGATTCCTGCTGTGGTCAAGATTCGTAATCTGT 960
Db 1021 ACCATTGGCGCAGCAGCTGTGATCGAAGATTCCTGCTGTGGTCAAGATTCGTAATCTGT 960
Qy 1081 AAGCTTCTCTATGAGAGTCAATGTTATCCGGGAAAGGTATAAAGAAATCCGGGAGAGAG 1140
Db 1081 AAGCTTCTCTATGAGAGTCAATGTTATCCGGGAAAGGTATAAAGAAATCCGGGAGAGAG 1140
Qy 1141 GTTGAGTGTATTTACGTCACAGAGGAGATTAAGAAAGTGAAGTCTGTTTGAAGAGTCA 1200
Db 1141 GTTGAGTGTATTTACGTCACAGAGGAGATTAAGAAAGTGAAGTCTGTTTGAAGAGTCA 1200
Qy 1201 CTTTCAAGAGAAGAAATCCGAGCAAAATGGTGAATTTGATCGCGAGTCAATGGAAGA 1260
Db 1201 CTTTCAAGAGAAGAAATCCGAGCAAAATGGTGAATTTGATCGCGAGTCAATGGAAGA 1260
Qy 1261 TCGGAGAAAGGAGAGCGGTAAAGAGCTGGAGGAGTTGCAATGATCTTAGCCAAATACA 1320
Db 1261 TCGGAGAAAGGAGAGCGGTAAAGAGCTGGAGGAGTTGCAATGATCTTAGCCAAATACA 1320
Qy 1321 GAGATCAACCAAGAAGAAGATCTATTGAGCTTCACTCTTACCAGCTACATTGATTGT 1380
Db 1321 GAGATCAACCAAGAAGAAGATCTATTGAGCTTCACTCTTACCAGCTACATTGATTGT 1380
Qy 1381 TACAGTGTAGTCTCTTGAAGGCTTATGTTAATGCCAGCTGAAACCAAGGCGGG 1440
Db 1381 TACAGTGTAGTCTCTTGAAGGCTTATGTTAATGCCAGCTGAAACCAAGGCGGG 1440
Qy 1441 ATAAATTTTGGTGTACGGTGTATGGAGTCAAGAGCAGCAGGAGTGTCTCAGTTTCA 1500
Db 1441 ATAAATTTTGGTGTACGGTGTATGGAGTCAAGAGCAGCAGGAGTGTCTCAGTTTCA 1500
Qy 1501 GCTCGAGGACCGAGTTTAGCAATCTTGCATATAAAACCGGATATGATGCTCCGGA 1560
Db 1501 GCTCGAGGACCGAGTTTAGCAATCTTGCATATAAAACCGGATATGATGCTCCGGA 1560
Qy 1561 GTCAATATCATTTGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTC 1620
Db 1561 GTCAATATCATTTGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTC 1620
Qy 1621 AGAAGATTAACTTCACTGTAATGTGAGGATCTCAATCTGTTGCTCAGATCTAGCGA 1680
Db 1621 AGAAGATTAACTTCACTGTAATGTGAGGATCTCAATCTGTTGCTCAGATCTAGCGA 1680
Qy 1681 ATCACTGTCTTATCCGCTGTGATACCCGAACTGGTCTCCAGCTCAATCAAAACCGCA 1740
Db 1681 ATCACTGTCTTATCCGCTGTGATACCCGAACTGGTCTCCAGCTCAATCAAAACCGCA 1740

Qy 1741 TTGATGACAAACAGCGGATTTGTACGATCGTCAAGGAAAGCGGATAAAGAGTGTAAACAA 1800
Db 1741 TTGATGACAAACAGCGGATTTGTACGATCGTCAAGGAAAGCGGATAAAGAGTGTAAACAA 1800
Qy 1801 CCAGCCGGTGTGTTGGATTTGAGGAGGAGGATGTAATCCGCAAAAGCGGATAAACCAG 1860
Db 1801 CCAGCCGGTGTGTTGGATTTGAGGAGGAGGATGTAATCCGCAAAAGCGGATAAACCAG 1860
Qy 1861 GGATTTGTTTACAAACATTTCAACAGTGGATTACATAACTTACCTTGCACCTTTGGATT 1920
Db 1861 GGATTTGTTTACAAACATTTCAACAGTGGATTACATAACTTACCTTGCACCTTTGGATT 1920
Qy 1921 ACAGATCAGATATTTTAGCAATCACTCATPAAGAACGTGAGCTGCAATGGAATATTGCGG 1980
Db 1921 ACAGATCAGATATTTTAGCAATCACTCATPAAGAACGTGAGCTGCAATGGAATATTGCGG 1980
Qy 1981 AAAAACCCGGTTTTAGTCTCAATTTACCTGATAGCCGCTGATTTTCAACCTGGCAAG 2040
Db 1981 AAAAACCCGGTTTTAGTCTCAATTTACCTGATAGCCGCTGATTTTCAACCTGGCAAG 2040
Qy 2041 ACTACGGAGATGATCAACAGCGGTGTCACTAACGTTGGGAGTCTTAACCTGATATACTCA 2100
Db 2041 ACTACGGAGATGATCAACAGCGGTGTCACTAACGTTGGGAGTCTTAACCTGATATACTCA 2100
Qy 2101 GTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTGTCAATCTTAAGAGACTTGTGTT 2160
Db 2101 GTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTGTCAATCTTAAGAGACTTGTGTT 2160
Qy 2161 AAACACCTGTGATCAGAGCTGAGCTATAGAGTATGGTTTGTATTGAGAAAGAAACAGA 2220
Db 2161 AAACACCTGTGATCAGAGCTGAGCTATAGAGTATGGTTTGTATTGAGAAAGAAACAGA 2220
Qy 2221 GGAGGAGGCTGGCTAGCTTTGACAGGCGGAGTGGCTGGCTCACTCTCAATCTG 2280
Db 2221 GGAGGAGGCTGGCTAGCTTTGACAGGCGGAGTGGCTGGCTCACTCTCAATCTG 2280
Qy 2281 ATGCAGCGAGTTAGAGTCCAAATCTCTGTAACTTGAAGACTAACTGA 2328
Db 2281 ATGCAGCGAGTTAGAGTCCAAATCTCTGTAACTTGAAGACTAACTGA 2328

RESULT 3

AAZ93813
ID AAZ93813 standard; DNA; 2492 BP.
XX AC AAZ93813;
XX DT 16-AUG-2000 (first entry)
XX DE Subtilase (Pigesp) of Solanum tuberosum.
XX SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomach; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;
KW crop protection; feed; foodstuffs; ss.
XX Solanum tuberosum.

Key	Location/Qualifiers
CDS	3..2492
FT	/*tag= a
FT	/product= Subtilase Pigesp

WO200022144-A2.

PD 20-APR-2000.

XX PF 12-OCT-1999; 99WO-EP07633.

XX PR 12-OCT-1998; 98EP-0119244.

XX PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
(PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

XX Berger D, Altmann T;
XX WPI; 2000-317995/27.
XX P-PSDB; AAY83303.
XX Novel recombinant DNA molecules encoding subtilisin-like serine
XX protease, useful for producing transgenic plants with altered stomata,
XX lower water consumption and enhanced diseased resistance
XX
XX Claim 1; Page 79-82; 101pp; English.
XX Sequences encoding SDI, a subtilisin-like serine protease, can be
XX used to produce transgenic plants with altered stomata
XX characteristics. These plants exhibit improved freshness,
XX increased dry weight, reduced leaf temperatures, reduced water loss
XX and lower water consumption and for enhancing the sugar and/or
XX protein content of plant leaves, modulating CO₂ uptake into and H₂O
XX release from leaves, for sustained photosynthesis under high
XX intensity conditions or for the improvement of disease resistance
XX of plants. The transgenic plants and cells of such plants are useful
XX in the preparation of feed, food or additives. This subfamily of
XX Solanum tuberosum is a homologue of SDI of Arabidopsis thaliana
XX (See AAZ93808).
XX
SQ Sequence 2492 Bp; 766 A; 480 C; 511 G; 734 T; 1 other;
Query Match 44.0%; Score 1024.6; DB 21; Length 2492;
Best Local Similarity 68.1%; Pred. No. 5.1e-309;
Matches 1455; Conservative 0; Mismatches 675; Indels 6; Gaps 2;
QY 189 AGAGGAGAGCCCTCTCTCGACTCTCTACTCTCTATGGCTCGGATTAAGAGATTGC 248
DB 356 AGGTGAAAATCGAGTTCTCGCTTTTGTTACTCTTACCATTGCAATTAAGAGTTTGC 415
QY 249 TGCTCAGTTGACTGAATCAGAAGCCGAGATPACTGAGATATTCACCTGAAGTTGTTGCACT 308
DB 416 AGCACTTCTATCGAAATGAGCTAAAGGCACTGAAGAAATCGAATATGTTATCAAT 475
QY 309 GAGACCTGACATGTTCTCAGTTCAACCACTTACTCTTACAAGTTCTTGGACTCGA 368
DB 476 ATATCCGGAGAGAAAGCTTGAGGTTCAAAACAATTATTTTACAAGTTCTTAGGACT--- 532
QY 369 CGGTTTGGAACTCCGGTGTATGGTCTAAATCTCGTTTGGTTCAGGCAATATATCG 428
DB 533 TAGTCTTACAAAGGAAGTACTTGGTTAAAGTCTGATTTGGTTCAGGCGGCAATATCG 592
QY 429 CGTCTTGTACTGGAGTTTGGCTGAAAGTCTAGCTTGTGATACCGGAATGCTTC 488
DB 593 AGTCTTGTACTGGAATTTGGCAGAAAGTCCAAAGTTTGTGATCATGGAATGTTCTCC 552
QY 489 GATTCCACGGAATGGAAGGATTTGCCAAGAGGAGAAAGTTTCAGTTCTTCGAGGTG 548
DB 653 TATTCCAAAGAAATGGAAGGTTTTCGCCAAGAGGAAAGAACTTCAATTTCTCAAGTTG 712
QY 549 TAACCGAAGCTAATCGGTGCTAGATTCTTCAATCAGAGGACCGGTGCTGCTAATTCACC 608
DB 713 CAATCGCAAGCTATTGTTGTCAGAGTTTTCAGATAGGACACATGATGCGATCAAGAC 772
QY 609 AGAGGATACCAAAACATGCTCGTGAATACATTTCCGGAAGGATTCACGCGGACCG 668
DB 773 ATCAAATCAATGATTTTATGAGGATTTATGATCACTTCGAGATTCTCAAGGCGCATGG 832
QY 669 GACTTCACACCGCTCAACAGTTGGTGGATCTCTGTTTGGATGGCGAATGTTCTTGGCAA 728
DB 833 TACACATACAGCATCTACTCAGGGGAGCTCCCGTTCCAAATGCGGAGTGTCTTGGAAA 892
QY 729 TGGAGCTGGTGTGGCTCGTGGATGGCTCTGAGGTGCATATGCAATGCAATGATGCTG 788
DB 893 TGGAGCAGGAGAGGCTCGAGGATGGCCCTGGTGTCTATGTCGATATACAAAGTTTG 952
QY 789 TTGTTTCAATGGTTGTTACAGTCTGACATTTAGAGCTATAGATGAGGATTCAGAA 848

DB 953 TTGCTCTAGTGGTTGTTATAGTTCTGATATACATTTCGAGCAATGGATAGCTATTAGAGA 1012
QY 849 TAAAGTCGATGTTCTTTCGCTTCCCTTGGCGGTTTCCCTATTCTTCTTGTATGATGACAC 908
DB 1013 TGGAGTAGACATATGTTCTCTTCAATTTGGTGGTTTCCCTGTTCCACTTTATAGGATAC 1072
QY 909 AATCGCCATTGGAACATTCGAGGCAATGGAACGCGTATATCTGTAATCTGTGAGCTGG 968
DB 1073 TATTGCTATTGGCAGTTTTCGAGCTATGGAACGTTGGAATTCAGTTATATGTTGCTGAGG 1132
QY 969 TAAACACGGTCCAAATCGAAAGCTCTGTTGCAACACAGCTCTTGGTCTCAACCATGG 1028
DB 1133 AAATAAGTGTCCAAATCTTAAGTTTCAGTAGCAAAATGAGGCTCTTGGATTCGCAATATGG 1192
QY 1029 CGCAGGACGCTGATCGAAGATTTCCGCTGTGGTCAGATTAGCAACGAAAGCTTCT 1088
DB 1193 TGCTAGCACATTTGACAGGAAATTTCCAGCAATATTTTCAGCTAGGTAAATGGCAATATGT 1252
QY 1089 CTATGGAGAGTCATTTGATCCGGAAAAGGTATTAAGAATGCCGGGAGAGAGGTTGAGGT 1148
DB 1253 GTATGGAGATCTTTCAGCCGGGCAACAAAGTTTCATAATTTCTCAGAAAAGTTCTTGAGAT 1312
QY 1149 GATTTAGCTACAGGAGAGATTAAGGAAGTGTCTTCTGTTGAGAGGTCACATTCGAAG 1208
DB 1313 TGTTTATCTCAATGACGGTGTATAATGGAAGTGAATTTTGCTTAAGAGGGTCTCTGCCAAG 1372
QY 1209 AGAAGAAATCCGAGGCAAAATGTTGATTGTCGCGGAGTCAATGGAAGATCGGAGAA 1268
DB 1373 AGCTAAAGTCCATGGAATAATTCGTTGTATGATCGTGGAGTTAATGGAAGAGCAGAGAA 1432
QY 1269 AGGAGAAGCGGTTAAAGAAAGCTGGAGGAGTTGCAATGATCTTAGCCAAATCAGAGATCAA 1328
DB 1433 AGGTCAAGTTGTTAAGAAATCAGTGGTGTGTCATGATCTTAGCAATCAGCAGTAAA 1492
QY 1329 CCAAGAGAAGATTTCTATGACGTTCACTCTTACCAGCTACATGATGATGTTGTTACATGA 1388
DB 1493 TATGAGGAAGATTTCTGTGGACGTACATGCTCTACCTGCAACATTTGTTGTTGACGA 1552
QY 1389 GTGAGTCTTCTGAGGCTTATGTTAATGCCAGCTGAAACCAAGCGCGGATTAATTTT 1448
DB 1553 ATCAATTCAGTTGCAAAAGCTATATGAATCAACGCGAAAACCAACAGCTCGAATCATATT 1612
QY 1449 TGGTGTACGTTGATTTGGAGGTCACGAGCACCGGAGTGGCTCAGTTTTCAGCTCCAGG 1508
DB 1613 TGGAGGAACAGTTATAGAAAATCTAGTGACCTGCTGTAGCAAAATTTTCTCAAGGG 1672
QY 1509 ACCGAGTTTACCCAAATCCTTCGATACATAAACCAGATATGATTCCTCCGGAGTCAATAT 1568
DB 1673 TCCAGTTTACTGATCCTTCAATTTCTCAAACTGATGATGATGCTCCAGGTGTCAACAT 1732
QY 1569 CATTCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTCTTATGATTCAGAAAGAT 1628
DB 1733 AATTGCTGTTGGCGGCAAAATCTAGGTCTAGTGGCTGGCTGAGGATTCAGAAAGAT 1792
QY 1629 TAATTCAGTATATGTCAGGACTTCAATGTTCTTCCACATGTTAGCGGATCACTGC 1688
DB 1793 AAATTCAGTCTCTTATCAGGAACTTCAATGGCTGTCCTCATGTAGTGGCAATGCTGC 1852
QY 1689 TCTTATCCGGTCTCATACCCGAATCTGCTCCAGCTGCAATCAAAATCCCATTTGATGAC 1748
DB 1853 ACTACTCCATTCATTTCTTAATGTTCCAGCTGCAATCAAAATCCCGCTTAATGAC 1912
QY 1749 AACAGCGGATTTGTCAGTCTCAAGGGAAGCGATTAAGGATGGTTAAACAAACCGCCGG 1808
DB 1913 AACTGCAGACACAAACCAACCAAGGAAACCAATCATGATGGTGTGACACAGCAGTGG 1972
QY 1809 TGTGTTTGGATTTGGAGCAGGCAATGTAATCCGCAAAAGCGGATTAACCCGGGATTTGGT 1868
DB 1973 ACTTTTCGCCATAGGAGCTGGACATGTAATCTCTGGAAGATCCGATGATCCCGGATTTGAT 2032
QY 1869 TTACAAACATTCAGAGTGGATTAATGATTAATCTGCTGACCTTCTGATTTCAAGATC 1928
DB 2033 ATATGACATTAATGCAATGACTATATACCTCACCTTTGCACTATTGTTTACAAAACTC 2092

QY 1929 AGATATTTAGCAATCACTATAGAACGTGAGTGCATATGAATATTCGGAAAAACCC 1988
 Db |||||
 QY 2093 TGAATCCTCAGCATTAACACAGAAATGTAGTCCAGACGTTTACAGAAAACAG 2152
 Db |||||
 QY 1989 GGGTTTATGCTCAATACCCGTCGATAGCCGTGATTTCAACAGTGGCAAGACTACGGA 2048
 Db |||||
 QY 2153 GGGTTTATGCTCAATACCCCTCTATTTCCGTAATCTTTAAGGCGAGGAAAACGAGAAA 2212
 Db |||||
 QY 2049 GATGATCACAAAGCGGTGCTACTAAAGTGGAGTCCCTACTCGATATCTCAGTGAATGT 2108
 Db |||||
 QY 2213 AATGATCACAAAGAGAGTGCATAATGTGGGAGTCCCTAATTCATCTCTCAGTTGAAT 2272
 Db |||||
 QY 2109 CAAGGCTCCAGAGGGGATCAAGATTTATGTCATCTCAAGAGACTTGTCTCAAAACAGCT 2168
 Db |||||
 QY 2273 TGTGGCACCAGAGAGGTAAAGTGAGAGTTAAACCGCAGCTCTGTTTAAACATGT 2332
 Db |||||
 QY 2169 GGAATCAGACCTGAGCTATAGATGATGTTGTTGATTTGAAGAAGAAAACAGAGAGGAA 2228
 Db |||||
 QY 2333 TAATCAAGTTTAAAGTTACAGAGTTTGGTTTATATCAAGGAAGAGA--ATTGGGACTCA 2389
 Db |||||
 QY 2229 GGTGGCTAGCTTGCACAGGGGAGTGAATGAGTGGGTCAGTCTCATATCTGATCGAGG 2288
 Db |||||
 QY 2390 AAGGAGAAGCTTTGCAGAGGACAAATGATGTGGATCAACTCCAGAGATAAATACAGAA 2449
 Db |||||
 QY 2289 AGTTAGAAGTCCAAATCTCTGTAACCTTGAAGACTAA 2324
 Db |||||
 QY 2450 AGTTAGAAGTCTTATTTAGTTCATGGGATCGGCATCAA 2485
 Db |||||

RESULT 4

AAZ93814
 ID AAZ93814 standard; DNA; 3140 BP.

XX AAZ93814;

XX 16-AUG-2000 (first entry)

XX Subtilase (St_P2Sca) of Solanum tuberosum.

XX SDD1; serine protease; subtilisin; transgenic plants; dry weight;
 KW stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;
 KW crop protection; feed; foodstuffs; ss.

XX Solanum tuberosum.

XX Key Location/Qualifiers
 FT CDS 1..2301
 FT /*tag= a
 FT /product= Subtilase St_P2Sca

XX W0200022144-A2.

XX 20-APR-2000.

XX 12-OCT-1999; 99WO-EP07633.

XX 12-OCT-1998; 98EP-0119244.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

XX Berger D, Altmann T;

XX WPI; 2000-317995/27.

XX P-PSDB; AAY83304.

XX Novel recombinant DNA molecules encoding subtilisin-like serine
 PT protease, useful for producing transgenic plants with altered stomata,
 PT lower water consumption and enhanced diseased resistance

XX Claim 1; Page 85-88; 101pp; English.

CC Sequences encoding SDD1, a subtilisin-like serine protease, can be
 CC used to produce transgenic plants with altered stomata
 CC characteristics. These plants exhibit improved freshness,
 CC increased dry weight, reduced leaf temperatures, reduced water loss
 CC and lower water consumption and for enhancing the sugar and/or
 CC protein content of plant leaves, modulating CO₂ uptake into and H₂O
 CC release from leaves, for sustained photosynthesis under high
 CC intensity conditions or for the improvement of disease resistance
 CC of plants. The transgenic plants and cells of such plants are useful
 CC in the preparation of feed, food or additives. This subfamily of
 CC Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
 CC (See AAZ93808).

XX
 SQ Sequence 3140 BP; 939 A; 627 C; 690 G; 882 T; 2 other;

Query Match 43.9%; Score 1020.2; DB 21; Length 3140;

Best Local Similarity 86.4%; Pred. No. 1.4e-307;
 Matches 1530; Conservative 1; Mismatches 754; Indels 18; Gaps 4;

QY 14 CTTTCTTCTCTCGCATATCTTTCTTCTATTTTGTCTTCTTCTCGTCAGAGATCCTCGAGA 73
 Db |||||
 QY 2 CTCATTTATTTCTCTTCTATGCTTTTACTATGTTTGTTCATACAGCTCAAGATT 61
 Db |||||
 QY 74 AGCAGCTTACATGTTTCAAGCTTCATCTTAATAGCGAAACCGCTAAACCTTTGCTCAA 133
 Db |||||
 QY 62 TGCAAACTTACATAGTTCAAGTTCATCCACATCGAGCAACAGACCCCTTTTAGCTCTA 121
 Db |||||
 QY 134 AGTTGATTCGCATCTTTCTTTCTCAAGAGCGGTTTGTAGTGTGAAGAGAGAGG 193
 Db |||||
 QY 122 AACTACAATGGCACCTTTCTTCTTTCGAAAAGCAGTTCTCT-----GGAGAAC 172
 Db |||||
 QY 194 AAGAGCCTTCTTCGACTTCTCTACTCTATGGCTCTGATGAGGATTTGCTGCTC 253
 Db |||||
 QY 173 AAGACTCGTCTTCTCGTCTTTTGTACTCTTACATCTCGATGGAAGTTTTCAGCTC 232
 Db |||||
 QY 254 AGTTGACTGAATCAGAAAGCGAGATCTAGATATTCACCTCAAGTTGTTGCACTGAGAC 313
 Db |||||
 QY 233 GACTCACTGAAGATGAGTTGAGTGTGTAAGGAATCTAATGATGTTGTCATACGTG 292
 Db |||||
 QY 314 CTGACCAATGTTCTCAGGTTCAAAACCACTTCTCTTACAAGTTCTTTGGGACTCGAGGTT 373
 Db |||||
 QY 293 CTGAGAGAGGCTTGAATTCAGACTACTTATTTCTTACAAGTTCTTTGGGATT---AAGTC 349
 Db |||||
 QY 374 TTGGAACTCCGGTGTATGGTCTAAATCTCGTTTGGTCAAGGCAATATCGGGTGC 433
 Db |||||
 QY 350 CAACGAGAGAGAGGCTTGGTTGAAGTCTGGATTGGTCAGGGGCGATCAITGGAGTGT 409
 Db |||||
 QY 434 TTGATCTGAGTTTGGCTTGAAGTCTTAGCTTTGACGATACCGGAATGCCCTTCGATTTC 493
 Db |||||
 QY 410 TGGNACTGAGTTTGGCCAGAAAGTCCAAAGTTTTCATGATCATGSGATGCCACCTGCTC 469
 Db |||||
 QY 494 CACGAAATGGAAGGGATTTCGCAAGAGGAGAAAGTTTCAGTTCTTCGAGCTGTAACC 553
 Db |||||
 QY 470 CACAGAAGTGGAGGGTGTCTGCCAAGGAGGACAGGATTTAATCTTCTAGTTGTAATC 529
 Db |||||
 QY 554 GGAAGCTAATCGTGTAGATTCTTTCATCAGAGGACACCGTGTCTGCTAATTCACAGAGG 613
 Db |||||
 QY 530 GCAAGCTTATTTGTCGAAGGTTTTCAGAAAAGGACATCTGTGGC---TTCAATGACAT 586
 Db |||||
 QY 614 AATCACCACCAATGCCTCGTGATACATTTCCGCAAGAGATTCAACGGGACACGGGACTC 673
 Db |||||
 QY 587 CATCACCAGATGCAAGTGGAGGAATATGTGTGCCACCGGATTTCCCATGCCATGGTACAC 646
 Db |||||
 QY 674 ACACCGCTCAACAGTTGGTGGATCCTCTGTTTCGATGGCGAATGTTCTTGGCAATGGAG 733
 Db |||||
 QY 647 ATACAGCATCCATGCTGGAGAGCTGCAAGTTCCATTTGGCTGTGTGCTCGGAATGGAG 706
 Db |||||
 QY 734 CTGTTGTGGCTGTGGGATGGCTCTCGAGCTCACATTCGAGTCTATATAAGTCTGTGGT 793
 Db |||||
 QY 707 CAGGGAGGCTCGAGGAGTGGCCCGGTCGCCACATTCGCAATATATAAGTATGCTGTT 766
 Db |||||
 QY 794 TCAATGTTGTTACAGCTCTGACATTTCTAGCAGCTATAGATGTAGCGATTCAAGATAAG 853
 Db |||||

Db 767 TCAGTGGTTGTTACAGCTCTGATATATCTTGCAGCAATGGATGTGGCCATCAGAGATGGAG 826
Qy 854 TCGATGTTCTTTTCGCTTTCCTTGGGGTTTCCTATTCCTTTGTATGATGACACAAATCG 913
Db 827 TAGACATATTGTCACTCTCACTTGGTGGCTTCCCTATTCCACTTTTATGATGATACATTG 886
Qy 914 CCATTGGAACATTCAGAGCCATGGAACGGGTATATCTGTAACTGTGACAGCTGGTAACA 973
Db 887 CCATTGGAAGTTTCCGAGCCATGAGCATGGAATTTCACTATATATGTCTGTCAGGGAATA 946
Qy 974 ACGGTCCAAATCGAAAGCTCTGTTGCAAAACACAGCTCCTTGGGTCTCAACCATTTGGCGCAG 1033
Db 947 ATGACCAATCCAAAGTTCACTAGTCCAAACCGTGTCTCTTGATGCGCATTTTGGTGTCTA 1006
Qy 1034 GCAGCTTGATCGAAGATTTCCCGCTGTGTGCTAGATTAGCCAAACGGAAGCTTCTCTATG 1093
Db 1007 GCACATTGACAGGAGATTTCCAGCGTCAGTTAGGCAACGGAAGTTTCTCTACG 1066
Qy 1094 GAGAGTCATTGTTATCCGGGAAGGTATAAAGAAATGCCGGGAGAGAGTTTGAGGTGATTT 1153
Db 1067 GAGAATCCTTGTATCCCTGGGAAGAAATTCCTAGCTCTCAGAAGAATCTTGAGATCGTTT 1126
Qy 1154 ACGTTCAGAGGAGATTAAGGAAGTCACTTCTGTTTGGAGGGTCACTTCCAAAGAGAAG 1213
Db 1127 ATGTAAAGGATAGGACAAGGAAGTGAATTTGCTTGAGAGATCGCTATCAAAAGCAC 1186
Qy 1214 AAATCCGAGCAAAATCGTGAATTTGTATCGCGAGTCAATGGAAGATCGGAGAAAGGAG 1273
Db 1187 AGTCCGAGGGAATATGGTTGTGTGTATAGGGAGTCAATGGAAGGCAAGAAAGGCC 1246
Qy 1274 AAGCGGTTAAGAGCTGGAGGAGTTGCAATGATCTTAGCCAAATACAGAGATCAACCAAG 1333
Db 1247 AGGTTGTGAAGGAGGAGGTGGTGTGCCATGATCTTAGCAATACAGCAATAAATATGG 1306
Qy 1334 AAGAGATTTATGACGTTCTCTCTACAGCTACATGATGTTGTTACATGAGTCAAG 1393
Db 1307 AGGAAGATTCATGATGTCATGTCTCCAGCAACGTTGATTGGCTTCGATGAATCAA 1366
Qy 1394 TCCTTCGAGGCTTATGTTAAATGCCACGGTGAACCAAGGCGCGGATATTTTGGTG 1453
Db 1367 TTCATTACAAACTCTGAACTCAACAAAAGACCAACAGCTCGATTCATATTTGGAG 1426
Qy 1454 GTACGCTGATGGAGGTACAGAGCACCGGAGGTGGCTCAGTTTTCAGCTTCGAGGACCGA 1513
Db 1427 GAACGGTAATAGGAAGTCTAGAGCACTGCAGTAGCTCAGTTTTCGTCAGGGGGCCAA 1486
Qy 1514 GTTTAGCCATCTCTGATATCTAAACCGGATATGATGCTCCGGAGTCAATATCAATG 1573
Db 1487 GCTATACGTATCTTCAATTTCTAAACCTGATTTGATTTGCTCCAGGGGTAAACATAATTG 1546
Qy 1574 CGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTTCAAGAGAGTTAACT 1633
Db 1547 CCGCTTGGCCACAAAACCTTAGGCCCCCAGTGTCTTCCGAAGATTCACGAAGAGTAAAT 1606
Qy 1634 TCAGTGAATGTAGGAACCTCAATGCTTGTCCACATGTTAGCGGAATCACTGTCCTTA 1693
Db 1607 TCAGTGTATGTAGGAGCTCAATGCACTGTCTCTATGATTTCAAGAGAGTTAACT 1666
Qy 1694 TCCGCTGTCATACCCGAATGCTCTCCAGCTGCAATCAATCCGATTTGATGACAAACAG 1753
Db 1667 TCCATTGAGCTATCTTAATGGACTCCAGCAGCAATAAGATCCGCAATTAATGACCACTG 1726
Qy 1754 CGAATTTGATAGTCTCAAGGGAAGCGGATAAGGATGGTAAACCAACCGCGGTGTGT 1813
Db 1727 CAGATACAGCTGATCATATGGGAAACCAATCATGATGAGATGACACCAAGCTAAACTTT 1786
Qy 1814 TTGCGATTGAGCAGGCGCATGTAATCCGCAAGCGGATAAACCAGGATGGTTTACA 1873
Db 1787 TTGAGCTGAGCTGGACAGCTGAACCTCGAAGAGCCATCGATCTCTGGATTTGATATG 1846
Qy 1874 ACATTCACACAGTGGATTACATAACTTACCTCTGCACTCTTGGATTCACAAAGATCAGATA 1933
Db 1847 ACATCCAGGTTGATGATATATCACTCATCTTTTGCACATCTTGCACATATCGGATACAGAAATCTGAG 1906

RESULT 5

AAZ93812

ID AAZ93812 standard; DNA; 3865 BP.

XX AAZ93812;

XX AC AAZ93812;

XX DT 16-AUG-2000 (first entry)

XX DE Subtilase (Pigesp) of Solanum tuberosum.

XX KW SDD; serine protease; subtilisin; transgenic plants; dry weight;

XX KW stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;

XX KW crop protection; feed; foodstuffs; ss.

XX OS Solanum tuberosum.

FH Key Location/Qualifiers

FT exon 3..551

FT /tag= a

FT /label= Exon 1

FT intron 552..986

FT /tag= b

FT /label= Intron 1

FT exon 967..1654

FT /tag= c

FT /label= Exon 2

FT intron 1655..1737

FT /tag= d

FT /label= Intron 2

FT exon 1738..2222

FT /tag= e

FT /label= Exon 3

FT intron 2223..2485

FT /tag= f

FT /label= Intron 3

FT exon 2486..3252

FT /tag= g

FT /label= Exon 4

XX WO200022144-A2.

XX PD 20-APR-2000.

XX

PF 12-OCT-1999; 99WO-EP07633.
XX
PR 12-OCT-1998; 98EP-0119244.
XX
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
FA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
XX Berger D, Altmann T;
XX
XX WPI; 2000-317995/27.
DR P-PSDB; AAY83302.
XX
XX Novel recombinant DNA molecules encoding subtilisin-like serine
PT protease, useful for producing transgenic plants with altered stomata,
PT lower water consumption and enhanced diseased resistance
XX
XX Claim 1; Page 72-76; 101pp; English.
XX
XX Sequences encoding SDD1, a subtilisin-like serine protease, can be
CC used to produce transgenic plants with altered stomata
CC characteristics. These plants exhibit improved freshness,
CC increased dry weight, reduced leaf temperatures, reduced water loss
CC and lower water consumption and for enhancing the sugar and/or
CC protein content of plant leaves, modulating CO₂ uptake into and H₂O
CC release from leaves, for sustained photosynthesis under high
CC intensity conditions or for the improvement of disease resistance
CC of plants. The transgenic plants and cells of such plants are useful
CC in the preparation of feed, food or additives. This subfamily of
CC Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
CC (See AA293808).
XX
XX Sequence 3865 BP; 1202 A; 707 C; 715 G; 1237 T; 4 other;
SQ
Query Match 25.9%; Score 602; DB 21; Length 3865;
Best Local Similarity 58.5%; Pred. No. 9.5e-177;
Matches 1344; Conservative 0; Mismatches 606; Indels 349; Gaps 3;
372 TTTTGGAACTCGGTGTATGCTAAATCTGGTTTGTCTAAGGACACAAATTATCGGGT 431
951 TTTTTCCTAATAGGTACTTGGTTAAAGTCTGGATTGTGCGGGCGCATCATTTGGAGT 1010
432 GCTTGATCTGGAAGTTGGCTCAAGTCTAGCTTTGACGATACCGGAATCCCTTCGAT 491
1011 TCTTGATCTGGAATTTGGCCAGAAAGTCAAGTTTGTGATCATGGAATGCTCCTAT 1070
492 TCCACGGAATGGAAGGATTTGCCAAGAAGGAGAAAGTTTCAGTTCTTCGAGCTGTAA 551
1071 TCCAAAGAAATGGAAGGTTCTGCCAAGAAGGAAAACTTCAATTTCAAGTTGCAA 1130
552 CGGAAGCTAATCGGTCTAGATTTCTCATCAGAGACACCGTGTGCTAATTCACGAGA 611
1131 TCGCAAGCTTATTTGGTCAAGGTTTTTCCAGATAGGACACATGATGGCATCAAGACATC 1190
612 GGAATCACCACAAATCGCTGTGAATACATTTCCGCAAGAGATTCACGGGACAGGGAC 671
1191 AAATCAATAGATTTATGGAGGATATGATATCACCTCGAGATTTCAAGGGCCATGGTAC 1250
672 TCACACCGCTCAACAGTTGGTGGATCTCTGTTTCGATGGGCAATGTTCTTGGCAATGG 731
1251 ACATACAGATCTACTGACGGGGAGGCTCCGTTCCAAATGGCGAGTGTCTTGGAAATGG 1310
732 AGCTGGTGTGGCTCGTGGGATGGCTCTGGAGTCAATTCGAGTCTATAAAGTCTGTGG 791
1311 AGCAGGAGAGGCTCGAGGATGGCCCTGGTGTCTATATCGCGATATACAAAGTTTGTGG 1370
792 GTTCAATGTTGTTTACAGTCTACATTTAGCAGCTATAGATGATGACCAATCAAGATAA 851
1371 GTCTAGTGGTGTATAGTTTCTGATATCTACTTGCAGCAATGGATGTAGCTATTAGAGATG 1430
852 AGTCGATGTTCTTTCGCTTTCCCTTGGCGGTTTCCCTATTCCCTTCTATGATGACACAA 911
1431 AGTAGACATATTTGCTCTTTCAATTTGGTGTGTTTCCCTGTTCCACTTTATGAGGATACTAT 1490

QY 912 CGCAATTTGGAACATTCGAGCCATGGAACGCGGTATATCTGTAAATCTGTGACGCTGGTAA 971
DB 1491 TGCTATTGGCAGTTTTCGAGCTATGGAACGCTGGAATTCAGTTATATGCTGCAGGAAA 1550
QY 972 CAACGGTCCAATCGAAAGCTCTGTTGCAACACAGCTCCTTGGGTCTCAACCAATTTGGCGC 1031
DB 1551 TAATGGTCCAATTTCTAAGTTTCAGTAGCAATGAGGCTCCTTGGATTGCCACTATTGGTGC 1610
QY 1032 AGGCAGCTTTGATCGAAGATTTTCCCGCTGT----- 1061
DB 1611 TAGCACACTTGACAGGAAATTTCCAGCAATTAATTCAGTAGGTATGTACATTTTGTTC 1670
QY 1062 -----GGTCAGA 1068
DB 1671 TTAATAATGATATTTCGGCTGTTTCCAGCCTAAATTAATGTGTCCTCATTTTCCCA 1730
QY 1069 TTAGCCAAACGAAAGCTTCTATGAGAGTCAATTTGATCCGGGAAAAGGTATAAAGAT 1128
DB 1731 ACAGGTAAATGGCAAGTATGTATGGAGAAATCTTGTACCCGGGCAACAAGTTCATAAT 1790
QY 1129 GCGGGAGAGAGTTGAGGTGATTACGTCACAGGAGGAGATAAAGGAACTGAGTTCTGT 1188
DB 1791 TCTCAGAAAGTTCTTGAGATGTTTATCTCAATGACGGTGATTAATGAAAGTGAATTTGC 1850
QY 1189 TTGAGAGGCTCACTTCCAGAGAGAAATCCGAGGCAAAATGTTGATTCGCGGA 1248
DB 1851 TTAAGAGGCTCTCTGCCAAGAGCTAAGTCCATGGAATAATCTGTATGATCGTGA 1910
QY 1249 GTCAATCGAAGATCGGAGAAAGGAGGCGTTTAAAGAACTGAGGAGTTGCAATGATC 1308
DB 1911 GTTAATGGAAGAGCAGAGAAAGGTCAAGTTGTTTAAAGAAATCAGGTGTTGTCATGATC 1970
QY 1309 TTAGCCAAATACAGAGATCAACCAAGAGAGATTTCTATTACGCTTCATCTTTACCAGCT 1368
DB 1971 CTAGCAAAATACAGAGTAAATATGAGGAGAAAGTCTGTGACGATCATGTCCTACCTGCA 2030
QY 1369 ACATTGATTTGGTTACACTGAGTCAGTCTCTCTCAAGCTTATGTTAATGCCACGGTGAAA 1428
DB 2031 ACATTGATTTGGTTGACGAATCAATTCAGTTCCAAAGCTATATGAACTCAACGCGAAA 2090
QY 1429 CCAAGGCGGGGATTAATTTTGGTGTACGGTGATTTGGGAGGTCAAGACACCGGAGGTG 1488
DB 2091 CCAACAGCTCGAATCATATTTTGGAGGAAACAGTTTATAGGAAAATCTAGTGCACCTGTGA 2150
QY 1489 GCTCAGTTTTCAGTCTCAGGACCGAGTTTAGCCAAATCTTTCGATATAAACCAGGATATG 1548
DB 2151 GCACAAATTTCTTTCAGGGGTCCAAGTTTACTGATCTTCAATTTCTCAAACCTGATGTG 2210
QY 1549 ATTGCTCCG----- 1557
DB 2211 ATTGCTCCAGTCAAGTTTATTGAACCAATTTATTATTAGATCATAGTAGCAAT 2270
QY 1558 ----- 1557
DB 2271 GTGACCAACAGGTTAGGGATTGAGCGCGTGAATCTATCATTTGATGCTGAATCAGGTA 2330
QY 1558 ----- 1557
DB 2331 GACTGCTTACATCACACCTTCCACAGACCATGATGAAAAACGGGATGCTCTTTTATAT 2390
QY 1558 ----- 1557
DB 2391 GCATGTGAAAAAACTTTTAATAAATAGTGTAAATGTTATGTTTGAACCTATATCTTCGTA 2450
QY 1558 -----GGAGTCAATATCATTTGGGCTTGGGCTC 1585
DB 2451 TAATCAGAAATCTTGAATCCGCCCTCTGCTCCAGGTGTCAACATAATTTGCTGTGGCGC 2510
QY 1586 AAAATCTAGGACCAACCGGACTTCTTTATGATTAATCAAGAGAGTAACTTCACGTGAATGT 1645
DB 2511 AAATCTAGGTCCTAGTGGCTCTGCTGAGGATTCAGAGAGTAACTTCACGTGCTTAT 2570
QY 1646 CAGGAATCTCAATGCTTTGTCACATGTTAGCGGAATCACTGCTCTTATCCGGTCTGCAT 1705

Db	878	GAGCGTTTTCGGCTATGGAGAGAGTGTTTTTGCTGTCTGCTGCTGTAATAGTGCTG	937
Qy	980	CAATCGAAAGCTCTGTGTGCAACACACAGCTCCTTGGGTCTCAACCAATGCGCGACGACGC	1039
Db	938	CTACTAGAGCTTCTGTGTGCAATGTGTCTCTTGGGTTATGACTGTGGTCTGGTACTT	997
Qy	1040	TTGATCGAAGATTTCCCGCTGTGTGTCAGATTAGCCAAACGGAAGCTCTCTATGAGAGT	1099
Db	998	TAGATAGAGATTTTCGGGCTTTTGGCAATCCGGTAACCGGAAACGACTTACCGGTGTTT	1057
Qy	1100	CAATTGATCCGGGAAAGGTATAAAGATATCCGGGAGAGAGGTTGAGGTGATTTACGTCA	1159
Db	1058	CCCTGTATAGCCGCTGTAGGAATGGGGA-----CGAAGCCGCTTGAATTGGTTTATAATA	1111
Qy	1160	CAGGAGGAGATAAAGGAAGTCAAGTTGAGAGGCTCACTTCCAAAGAGAAGAAATCC	1219
Db	1112	AAGGAAATAGTAGTTTCGAGTAATCTTGTGTACCTGGTTCGCTGATTCGATGATGTTTC	1171
Qy	1220	GAGGCAAAATGGTGATTTGTGATCCGGAGTCAATGGAAGATCGGAAAGGAGAAGCGG	1279
Db	1172	GTGGGAAGATTTGTTTGTGATAGAGGTGTAATGCTAGAGTTGAGAAAGGAGCTCTGG	1231
Qy	1280	TTAAAGAGCTGGAGAGCTTGCAATGATCTTAGCCAATACAGAGATCAACCAAGAAAG	1339
Db	1232	TTAGAGATGCTGTGTGGTTTGGGATGATPAATGGGCAATACTGCTCGAGTGGAGAGGAGC	1291
Qy	1340	ATTCTATTGACGTTTCATCTCTTACCAGCTACATTTGATTGTTTCACTGAGTCAGTCCCTC	1399
Db	1292	TTGTGGCGGATAGTCAATTTGTTCCCGGATCGCTGTAGGAAGAAGTGGTGATTTAC	1351
Qy	1400	TGAAGCTTATGTTTAATGCCACGGTGAACCAAGGCGCGGATAAATTTTGGTGGTACGG	1459
Db	1352	TTAGGAGATGTTTAAGTCAGATCTTAAACCAACCGCTCTCTTTGTTTTAAAGGAACGG	1411
Qy	1460	TGATTTGGAGGTCACGAGCACGGAGGTGGCTCAGTTTTCAGCTCGAGGACCGAGTTTAG	1519
Db	1412	TTCTTGACGTTTAAGCCGCTCTCCTGTGGTGTGCTGTTTTAGCTCGAGAGGTCCTAACTAG	1471
Qy	1520	CCAATCCCTTCGATACTAAACCGGATGATGTTGCTCGGGAGTCAATATCATTTCCGGCTT	1579
Db	1472	TTACTCTGAAATCTTGAGCCCTGATGTTATGTGCTGTGAGTTAAATTTTGGCTGGTT	1531
Qy	1580	GGCTCAAAATCTAGGACCAACCGGACTTCCTTATGATTCAAGAAGAGTTAACTTCACTG	1639
Db	1532	GGTCTGACGCTATTGGTCTACTGTTTGACAGGACTCTAGGAGACTCAGTTCAACA	1591
Qy	1640	TAATGTGAGGAACCTTCAATGTTCTGTCCACATGTTAGCGGAATCACATGCTCTTTATCCGGT	1699
Db	1592	TCATGTGAGGTACGTCAATGTCTATGCCCCACACATCAGTGGTTTTAGCGGCTCTTTGAAAG	1651
Qy	1700	CTGCATACCGAACTGTGCTCCAGCTCGAATCAATCCGATTTGATGACACAGCCGATT	1759
Db	1652	CAGTCACTCCCTGATGGAGTCGAGTGCTATCAAAATCAGTCTCTCACTACAGCTTACG	1711
Qy	1760	TGTACGATCGTCAAGGGAAGCGATAAGGATGGT-----AACAAACACAGCCGGTGTGT	1813
Db	1712	TTCTTGACAACACCAACGCTCCTCTCCATGATGCTGCAGACACAGCCTATCTAACCCAT	1771
Qy	1814	TTGGGATTGGACGAGGCGATGTGAATCCGCAAAAGGCGATAAACCCGGGATTTGGTTACA	1873
Db	1772	ATGCTCACGGCTCGGGCCATGTAGATCCCAAAAGGCTCTCTCACCGAGTCTTGTCTACG	1831
Qy	1874	ACATTCAACAGGTGATTTACATAACTTACCTCTGCACTCTTGATTTACAGATCAGATA	1933
Db	1832	ACATCTCAACCGAGGAATACATCAGGTTTTTGTGCTCTCTAGACTACAGATCGATCACA	1891
Qy	1934	TTTTAGCAATCACTATAAGAAAGTGAAGTGAATGGAATATTGCGGAAAAAACCCGGGTT	1993
Db	1892	TTGTTGGAT-----TGTGAAGGACCTAGCGTTAACTGCTCGAAGAGATTTCTCAGATC	1945
Qy	1994	TTAGTCTCAATTTACCCGTCGATACCGGTGATTTTCAACGTGGCAAGACTACGGAGATGA	2053

Db	1946	CTGGTCAGCTCAACTATCCCAAGTTTCTCGGTTTGTGTTGGGGGTAAAGAGTTGTGCGGT	2000
Qy	2054	TCACAAGCGGTGTCCTACTAAAGTTGGGAGTCCTAACTCGATATACTCAGTGAATGTCAAGG	2113
Db	2066	ACACTCGGGAAGTAAACAATGTTGGTGCAGCAAGCTCGGTTTACAAAGTCACGGTTAATG	2065
Qy	2114	CTCCAGAGGGGATCAAAAGTTATTCTCAATCCTAAGAGACTTGTGTTTCAAAACAGTGGATC	2173
Db	2066	GAGCTCCTAGTGTGGATCTCTGTTAAACATCATGAACCTTCGTTTAAAGCGTGGAG	2125
Qy	2174	AGACGCTGAGCTATAGAGTAGTGTGTTGTAATTGAAGAAAGAAAAACAGAGGAGGAAGGTGG	2233
Db	2126	AGAAAGAGGCTACACAGTCACGTCACGTTTGTAGCAAGAAAGGAGTCAGTATCAGCAACAAG	2185
Qy	2234	CTAGCTTTG 2242	
Db	2186	CTGAGTTTG 2194	
RESULT 7			
ABZ13260			
ID	ABZ13260 standard; DNA; 2265 BP.		
XX	AC ABZ13260;		
XX	AC		
DT	21-JAN-2003 (first entry)		
XX	XX		
DE	Arabidopsis thaliana stress regulated gene SEQ ID NO 1065.		
XX	XX		
KW	Arabidopsis thaliana; plant; gene; stress; transgenic; ds.		
XX	XX		
OS	Arabidopsis thaliana.		
XX	XX		
PN	WO200216655-A2.		
XX	XX		
PD	28-FEB-2002.		
XX	XX		
PF	24-AUG-2001; 2001WO-US26685.		
XX	XX		
PR	24-AUG-2000; 2000US-227866P.		
PR	26-JAN-2001; 2001US-264647P.		
PR	22-JUN-2001; 2001US-300111P.		
XX	XX		
PA	(SCRI) SCRIPPS RES INST.		
PA	(SYGN) SYNGENTA PARTICIPATIONS AG.		
PI	Harper JF, Kreps J, Wang X, Zhu T;		
XX	XX		
DR	WPI; 2002-304127/34.		
XX	XX		
PT	Identifying a stress condition to which a plant cell has been exposed		
PT	and producing plants with increased tolerance to these abiotic stresses		
PT	-		
XX	Claim 144; SEQ ID NO 1065; 577pp + Sequence Listing; English.		
XX	XX		
CC	The invention relates to identifying a stress condition to which a plant		
CC	cell has been exposed, comprising:		
CC	(a) contacting nucleic acid representative of expressed polynucleotides		
CC	in the plant cell with an array or probes representative of the plant		
CC	cell genome; and		
CC	(b) detecting a profile of expressed polynucleotides in the plant cell		
CC	characteristic of a stress response. The method is useful in the		
CC	production of transgenic plants, cells and seeds and in producing plants		
CC	with increased tolerance to abiotic stress. The present sequence is that		
CC	of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used		
CC	in methods of the invention.		
CC	Note: The sequence data for this patent is not represented in the printed		
CC	specification but is based on sequence information supplied to Derwent by		
CC	the European Patent Office.		
XX	Sequence 2265 BP; 519 A; 510 C; 542 G; 694 T; 0 other;		
SQ	XX		

Query Match	20.7%	Score 481.8	DB 24	Length 2265
Best Local Similarity	55.3%	Pred. No. 2.7e-139		
Matches 1033	Conservative 0	Mismatches 812	Indels 24	Gaps
Qy	380	ACTCCGGTGTATGGTCTAAATCTCCGGTTTCGGTCAAGGCACAAATATTCGGCGTCTCTTGATA	439	
Db	344	AATTCGGTGTTCAGGATCTCGGTTCCTTCTCTAACGGCGTATATCTCGAGATTTTAGATA	403	
Qy	440	CTGAGATTTGGCCTGAAAGTCTTAGCTTTTGACGATACCGGAATGCCCTTCGATTCACCGGA	499	
Db	404	CTGGCTTTGGCCTGAATCTAGAAGCTTCGATGATCTGATATGCCCTGAGATCCCTTCTTA	463	
Qy	500	AATGGAAGGGATTTGCCAAGAAGGAAAGTTTCAGTCTCTCGAGCTGTAAACGGGAAGC	559	
Db	464	AATGGAAGGAGAAATGTGAATCTGGTTCCGATTTCCGATTTCCAAAGTTGTGTAAACAAGAC	523	
Qy	560	TAATCGGTGTAGATTTCTTCATCAGAGGACACCGTGTCCGCTAAATTCACACAGAGGAATCAC	619	
Db	524	TTATCGGAGCTAGAGCTTCTCCAAAGGATTTCAAATGGCTTCTGGTGGTGGTTTTCGA	583	
Qy	620	CAAAATGCTCTCGTGAATACATTTCCGCAAGAGATTCACGGGACACGGGACTCACACCG	679	
Db	584	GTAA-----GCGTGAATCTGTTTCTCTCGTGATGTTGACGGACATGGAACACATACTT	637	
Qy	680	CCTCAACAGTTGTGGATACCTCTGTGTTTCGATGCGGGAATGTTCTTTGGCAATGGAGCTGGTG	739	
Db	638	CAACTACCGCGCGGATCCGCGCTTAGAAACGCTAGCTTTCCTCGGTACGCGCCGGTA	697	
Qy	740	TGGCTCTGGGATGGCTCCTGGAGCTCACATTCGAGCTATAAGCTGTGTTGTTCAATG	799	
Db	698	CGGCCAGAGGTATGGCCACTCGTCTCGTGTGCTACTTATAAAGTTGTGTGGAGTACTG	757	
Qy	800	GTGTGTTACAGCTCTGACATTTCTAGCAGCTATAGATGTAGCGATTCAGATAAAAGTCCGATG	859	
Db	758	GTGTGTTTGGATCTGATATACTAGTCTGCTATGGATCGAGCTATACITGTATGGTGTGAUG	817	
Qy	860	TTCTTTCCGCTTTCCCTTTGGCGGTTTCCCTATCTCTTTGTATGATGACACAATCGCCATTG	919	
Db	818	TGCTTTCTGTTATCTCTTGGTGGTGTCTGCTCCGTTATTCGCGATACGATTTCCGATTG	877	
Qy	920	GAACATTTCCGAGCCATGGACGGGTATATCTGTAACTCTGCGAGCTGGTAAACAACGGTC	979	
Db	878	GAGCGTTTTCGGCTATGGAGAGAGGTGTTTGTGTCTTGTCTCTGCTGGTAAATAGTGGTC	937	
Qy	980	CAATTCGAAAGCTCTGTTTGCAAAACACAGCTCTCTTGGGTCTCAACATTTGGCGCAGGCACGC	1039	
Db	938	CTACTAGAGCTTCTGTTGCCAATGTTGCTCTTGGGTATGACTGTGTGGTCTGTACTT	997	
Qy	1040	TTGATCGAAGATTTCCCGTGTGGTCAAGATTAGCAACCGAAAGCTTCTCTATGGAGAGT	1099	
Db	998	TAGATAGAGATTTTCCGGCTTTTCGAACTCTCGGTAACGGGAAACGACTTTACCGGTGTTT	1057	
Qy	1100	CATTGTTATCCGGAAAGGTATAAAGAAATGCCGGAGAGAGTTGAGTGCATTTACGTCA	1159	
Db	1058	CGCTGTATAGCGGTGTAGGAATGGGA-----CGAAGCCGCTTGAATTCGTTTATAATA	1111	
Qy	1160	CAGGAGGAGATAAAGGAAGTGAAGTCTCTGTTTACAGGGTCACTTTCCAAAGAGAAGAAATCC	1219	
Db	1112	AAGGAATAGTGTTCAGTAACTTTGTTTACCTGTTCCGTTGATTCAGATATTGTTTC	1171	
Qy	1220	GAGCAAAATGGTGATTTGTGATCCGGAGTCAATGGAAGATCGGAAAGGAGGAAGCGG	1279	
Db	1172	GTGGGAAGATTGTTGTTGTGATAGAGGTGTAACTGCTAGAGTTTGAGAAAGGAGCTGTGG	1231	
Qy	1280	TTAAGAAGCTGGAGGAGTTGCAATGATCTTAGCCATACAGATCAACCAAGAAGAAG	1339	
Db	1232	TTAGAGATGCTGTGGTGTTAGGGATGATAATGGCGAATCTGCTGCGAGTGGAGAGGAGC	1291	
Qy	1340	ATTCTATTGAGTTCACTCTTACAGCTACATTTGATTTGGTTACCTGACCTGAGTCCCTTC	1399	
Db	1292	TTGTGGCGGATAGTCAATTGCTTCCCGGATCGCTGTAGGAAGAGACTGTTGATTTAC	1351	
Qy	1400	TGAAGGCTTATGTTTAATGCCACGGTGAACCAAAAGGCGCGGATAATTTTGGTGGTACGG	1459	

D	b	1352	TTAGGAGTATGTTAACTCAGATTCTTAACCAACCGCTCTTCTGTGTTTTAAGGAACGG	1411
Q	y	1460	TGATTGGGAGGTCA CGAGCACCGAGGTGGCTCAGTTTTTCAGCTCGAGGACCGAGTTTAG	1519
D	b	1412	TTCTTGACGTTAAGCCGCTCTCTGTGTTGCTGCTTTTAGCTCGAGAGGTCCTAATACTG	1471
Q	y	1520	CCAATCTCTCGATACTAAACAACCGGATATGATTGCTCCGGAGTCAATATCATTTGCCGCIT	1579
D	b	1472	TTACTCTCGTAAATCTTTGAAGCCCTGATGTTATTGGTCTCGAGTTAATATTTTGGCTGGTT	1531
Q	y	1580	GGCTCAAAATCTAGGACCACAAACCGGACTTCCTTATGATTCAAGAAGAGTTAACTTCAC TG	1639
D	b	1532	GGTCTGACGCTATTTGTTCTCTACTGTTCTTGACAAGACTCTAGGAGGACTCAGTTCAACA	1591
Q	y	1640	TAATGTGAGAACTTCAAATGTTCTTTGCCATGTTTAGCGGAATCACTGCTCTTTATCCGGT	1699
D	b	1592	TCATGTCAGTACGTCATATGTCATGCCACACATCAGTGGTTTTAGCGGCTCTTTTGAAG	1651
Q	y	1700	CTCGATACCGGAAGTCTGCTCAGCTGCGAATCAAAATCCGATTTGATGACACAAGCCGATT	1759
D	b	1652	CAGCTCACCTCTAGTGGAGTCCGAGTGSCTATCAAAATCAGCTCTCATGACTACAGCTTACG	1711
Q	y	1760	TGTACGATCGTCAAGGGAAGCGATAAAGATGGT-----AACAAACCAAGCCGGTGTGT	1813
D	b	1712	TTCTTGACAACCAACAGCTCTCTCCATGATGCTGCGACACAACAGCCTATCTAACCCAT	1771
Q	y	1814	TTGCGATTGGAGCAGGCGCATGTGAATCCGCAAAAGCGATAAACCCGGGATTTGGTTTACA	1873
D	b	1772	ATGCTCAGCGCTCGGCCCATTGATATCCCCAAAAGGCTCTCTCACCAAGTCTTGTCTCAG	1831
Q	y	1874	ACATTCACACAGTGGATTTACATACTTTACCTCTGCACTCTTTGGATTTCACAGATCAGATA	1933
D	b	1832	ACATCTCAACCGGAGGAATACATCAGGTTTTTTGTGCTCTCTAGACTACACAGTCCGATCACA	1891
Q	y	1934	TTTTAGCAATCACTATAGAACGTCGAGCTGCAATGGAATATTGCGAAAAACCCGGGTT	1993
D	b	1892	TTGTTGGAT-----TGTGAACCGACCTAGCGTTAACTGCTCGAAGAAGTTCAGATC	1945
Q	y	1994	TTAGTCTCAATTACCGCTCGATAGCCGTGATTTTCAAACTGGCAAGACTACGGAGATGA	2053
D	b	1946	CTGGTCAGCTCACTACCCCAAGTTTCTCGTTTTTTGTTGGGGTAAAGAGTTTGTGCGGT	2005
Q	y	2054	TCACAAGCGGTGCACTAACGTTGGGAGTCTTAATCGATATACTCAGTGAATGTCAAGG	2113
D	b	2006	ACACTCGGGAAGTAACAAATGTTGGTGCAGCAAGCTCGGTTTACAAAGTGACGGTTAATG	2065
Q	y	2114	CTCCAGGCGGATCAAGATTATTGTCAACTCAAGAGACTTGTGTTCAACACGTGGATC	2173
D	b	2066	GAGCTCCTAGTTCGGGAATCTGTGTTAAACCATCGAAACTTCGTTTTAAAGCGTGGAG	2125
Q	y	2174	AGACGCTGACTATAGATATGTTTTGTATTGAAGAAGAAAAACAGAGGAGGGAAGTGG	2233
D	b	2126	AGAGAAGAGGTACACAGTCACGCTTTGTTAGCAAGAAAGGAGTGAGTATGACCAACAGG	2185
Q	y	2234	CTAGCTTTG 2242	
D	b	2186	CTGAGTTTG 2194	

RESULT 8
AAC47609
ID AAC47609 standard; DNA; 2434 BP.

XX AC AAC47609;
XX AC
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54454.
XX KW Hybridisation assay; generic mapping; gene expression control;
XX KW protein identification; signal transduction pathway;
XX KW metabolic pathway; promoter; termination sequence; ss.

RESULT 8	
AAC47609	
ID	AAC47609 standard; DNA; 2434 BP.
XX	
XX	
AC	AAC47609;
XX	
XX	
DT	18-OCT-2000 (first entry)
XX	
XX	
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 54454.
XX	
KW	Hybridisation assay; genetic mapping; gene expression control;
KW	protein identification; signal transduction pathway;
KW	metabolic pathway; promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX XX
PD 06-SEP-2000.
XX PF
PF 25-FEB-2000; 2000EP-0301439.
XX XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128034.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149436.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

Db	1731	ATCCAAGGAGGTTAAGTTCATATATACTGCTGGAACCTTCAATGTGTCATGTAA	1790	PR	23-APR-1999;	99US-0130510.
Qy	1676	GCGGAATCACTGCTTTATCGGTTCTGCATACCGGAACCTGCTCCAGCTGCAATCAAT	1735	PR	23-APR-1999;	99US-0130891.
Db	1791	GTGTTAGTCTTGTGATCAATCAAGGCATCCAGATTGGATCTCGCAGCAATCAAT	1850	PR	30-APR-1999;	99US-0132048.
Qy	1736	CCGCAITGATGACAACAGCGGATTTGATCGTCAAGGGAAGGATTAAGGATGGTA	1795	PR	30-APR-1999;	99US-0132407.
Db	1851	CAGTCTCATGACAACTGCTTATGTTTCATGACAACTGTTTAAAGCCTCTTACGGATGCAT	1910	PR	04-MAY-1999;	99US-0132484.
Qy	1796	ACAAACAGCGGT-----GTGTTGCGATTGGAGCAGGCGATGCAATCCGCCAAAGG	1849	PR	05-MAY-1999;	99US-0132485.
Db	1911	CAGGAGCAGTCTCTTCATCGCTTATGATCAAGTGCAGGACATATAGATCCTTTAAGAG	1970	PR	06-MAY-1999;	99US-0132486.
Qy	1850	CGATAAACCGGGATGTTTACAACTTCAACAGTGGATTACATAACTTACCTCTGCA	1909	PR	06-MAY-1999;	99US-0132487.
Db	1971	CTACAGATCTGTTGGTCTACGACATTTGGAATTTTGAATTCCTCTGCA	2030	PR	07-MAY-1999;	99US-0132863.
Qy	1910	CTCTTGGATTCACAAGATCAGATATTTTGAATCACTCATTAAGAAGTGAAGTGAATG	1969	PR	11-MAY-1999;	99US-0134256.
Db	2031	CTCAAGATTTAAGTCCATCAGCTTAAGGTATTCAAAAACATTCAAAACAGAACTGCA	2090	PR	14-MAY-1999;	99US-0134258.
Qy	1970	GAATATTGGGAAAAACCGGGTTTGTAGTCTCAATACCGTGCATAGCGGTGATTTTCA	2029	PR	14-MAY-1999;	99US-0134219.
Db	2091	AACACACTCTTGGCAAGATCCGGAACTTGAACCTACCGCGGATATCAGCTTTTCTCC	2150	PR	14-MAY-1999;	99US-0134221.
Qy	2030	AACGTGCA-----AGACTAGGAGATGATCAAGGCGTGCATCAAGTTGGGAGTC	2083	PR	14-MAY-1999;	99US-0134218.
Db	2151	CAGAAACACACATGTTAAAGCTATGACCTTTAAGAAGACAGTGACCAATGTTGTCCTC	2210	PR	14-MAY-1999;	99US-0134221.
Qy	2084	CTAACTCGATATCACTAGTGAATGTCAGGCTCCAGAGGGGATCAAAAGTTATTGTCAATC	2143	PR	18-MAY-1999;	99US-0134768.
Db	2211	ACATTTCAAGTCAAGGTTTCTGTCGCTCCGCAATCAAGGCGCATCGTAACTGTCAGC	2270	PR	18-MAY-1999;	99US-0134370.
Qy	2144	CTAAGACATTTGTTTCAACACAGTGGATCAGACGCTGACGCTGACCTATAGAT	2192	PR	18-MAY-1999;	99US-0134724.
Db	2271	CCAAAACACTCACTTCACTTCAAGGACCAAGGCTTTCTCTACACGGT	2319	PR	08-JUN-1999;	99US-0138094.
RESULT 9						
AC	AAC46158					
XX	ID	AAC46158 standard; DNA; 2637 BP.				
XX	AC	AAC46158;				
XX	DT	18-OCT-2000 (first entry)				
XX	DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 49118.				
XX	KW	Hybridisation assay; genetic mapping; gene expression control;				
XX	KW	protein identification; signal transduction pathway;				
XX	KW	metabolic pathway; promoter; termination sequence; ss.				
XX	OS	Arabidopsis thaliana.				
XX	PN	EP1033405-A2.				
XX	PD	06-SEP-2000.				
XX	PF	25-FEB-2000; 2000EP-0301439.				
XX	PR	25-FEB-1999; 99US-0121825.				
XX	PR	05-MAR-1999; 99US-0123180.				
XX	PR	09-MAR-1999; 99US-0123548.				
XX	PR	23-MAR-1999; 99US-0125788.				
XX	PR	25-MAR-1999; 99US-0126264.				
XX	PR	29-MAR-1999; 99US-0126785.				
XX	PR	01-APR-1999; 99US-0127462.				
XX	PR	06-APR-1999; 99US-0128234.				
XX	PR	16-APR-1999; 99US-0128714.				
XX	PR	16-APR-1999; 99US-0129845.				
XX	PR	19-APR-1999; 99US-0130077.				
XX	PR	21-APR-1999; 99US-0130449.				

PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147535.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149829.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	26-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.

PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match

18.2%; Score 424.4; DB 21; Length 2637;

Best Local Similarity 52.6%; Pred. No. 2.6e-121; Indels 51; Gaps 5;

Matches 1075; Conservative 0; Mismatches 916;

QY	200	CTTCTTCTCGACTTCTCTACTCTCTATGGCTCTCGGATTGAAGGATTGCTGCTCAGTTGA	259
Db	324	CTCTCTCTCCATATATCCACATACACACCGTTTCCATGCTTCTCCGCTAGACTCA	363
QY	260	CTGAATCAGAAGCGAGATACCTGAGATATTACCTGAAGTTGTGCACTGAGACCTGAAC	319
Db	384	CTTCACAAGCGCTAGTCAACTCTTAGACCATCTCATGTCATCTCTGTTATCCCGAAC	443
QY	320	ATGTTCTTAGTTTCAAAACCACTTACTCTTACAGTTCTTGGGACTCGACGGTTTGGAA	379
Db	444	AAGTCCTGCTCACTTGACACCACTCGTTCCCTGAGTTTCTTGCTTCTAGTCCACCGACA	503
QY	380	ACTCGGTGTATGCTTAATCTCGGTTTGGTCAAGSCAAATTTATCGGGTGTCTTGATA	439
Db	504	AAGCGGTCTACTCGAAGAGTCTGATTTCCGGTTCGGATCTAGTTATCGGAGTTATCGATA	563
QY	440	CTGGAGTTTGGCTTGAAGTCTTAGCTTTGACGATACCGGAATGCCCTTCGATTCCACGGA	499
Db	564	CTGTTGTTTGGCCGGAAGACCTAGCTTTGATGACCGTGTCTTGTCCTGTTCCCATTA	623
QY	500	AATGGAAGGATTGCGCAAGAGGAAAGTTTCAGTTCTTCGAGCTGTAAACGGHAGC	559
Db	624	AATGGAAGGCAATGTATCGCTTCCCAAGATTTTCCGGAGTCTGCTTGTAAACGTAAC	683
QY	560	TAATCGGTGTAGATTCTTCAATCAGAGACACCGTGTGCTTAATTCACAGAGAAATCAC	619
Db	684	TCGTCCGAGCTAGATTCTTCTCGGGTGGTA-----TGAAGCAACCAACGGGAAAA	734
QY	620	CAAAATGCTCGTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCG	679
Db	735	TGATGAACGACTGAGTTTCGCTCTCCGGTGNACTCCGATGACATCGGACTCACACAG	794
QY	680	CCTCAACAGTTTGGTGGATCTCTGTTTCGATGGCGAATGTTCTTGGCAATGGAGCTGGTG	739
Db	795	CTTCAATCTCCGCCGCCGTTAGCTTTTCCGGCGTCAACTCTCGGCTACGCTCACGGCG	854
QY	740	TGGCTCGTGGATGGCTCTCGGAGCTCACATTGAGTCTATAAAGTCTGTGGTTCATG	799
Db	855	TCGCTGCTGGATGGCTCCAAAAGCTAGACTCGCGGCTTACAAAGTCTGTGGAACTCCG	914
QY	800	GTTGTTACAGCTCGACATTTCTAGCAGCTATAGATGAGGATTCAAGATAAAGTCGATG	859
Db	915	GTTGTTACAGCTCGATATCTTAGCCGCTTTTCGACACCGCGGTTGCCGAGGTTCCATG	974
QY	860	TTCCTTTCGCTTTCCTTGGCGGTTTCCCTATTCTCTTGTATGATGACACAAATCCCATG	919

Db 975 TCATCTCTCTCCGTTGGAGGCGTCGTGGTTCCTTATTACCTAGACGGCTATCGCTATTG 1034
Qy 920 GAACATTCAGAGCAATGGAACCGGGTATATCTGTAATCTGTGAGCTGTGTAACAACGGTC 979
Db 1035 GAGCTTTGGAGCTATTGACAGAGGCATATCGTCTCTGCTCCGCGGTAAACGGAGTC 1094
Qy 980 CAATCGAAAGCTCTGTTGCAAAACAGAGCTCTTGGGTCTCAACCATTTGGCGGAGGACGC 1039
Db 1095 CCGGTGCTTTAAGCGGTGACAAATGTTGCTCCGTGGATGACAAACAGTCGGAGCTGGAACAA 1154
Qy 1040 TTGATCGAAGATTTCCGCTGTGTCAGATTAGCCACGGAAGCTTCTCTATGGAGGT 1099
Db 1155 TCGATAGGGAATTTCCGCGCAATGTGAACCTCGTAACGGGAAGATGATTTCTTGGTGTTA 1214
Qy 1100 CATTTGATCCGGGAAGAGTATAAA-----GAATGCCGGGAGAGAGGTTGAGGTGA 1150
Db 1215 GTGTATACGGTGGACCGGGTCTGATCCGGTCTGAATGATGATACCGCTTGTATTACGGTGT 1274
Qy 1151 TTTACGTCACAGGAGGAGATAAAGGAAGTGTCTGTTGAGAGGTCACTTCCAAAG 1210
Db 1275 GTTACTAGCGGCGATGGTTACTCATCGTCTCTGTCTGTGAAGCTCGTTGGATCCGA 1334
Qy 1211 AAGMAATCCAGGCAAAATGATGTTGATCGCGGAGTCAATGGAAGATCGGAAAG 1270
Db 1335 ATTAGTGAAGGGAAGATCGTTCTTTGTGATAGAGGAATCAATCTAGAGCAACCAAG 1394
Qy 1271 GAGAAGCGGTTAAGAGCTGGAGGTTGCAATGATCTTAGCAATACAGAGATCAACC 1330
Db 1395 GTGAGATCGTACGGAATAATGGAGGCTTGGGATGATTAAGCAATGGTGTGTTGAGC 1454
Qy 1331 AAGAAGAAGATCTTATTGAGGTTCACTCTTACCAGCTACATTAATGGTTTACACTGAGT 1390
Db 1455 GTGAGGTTTAGTAGCTGATGCAAGTGTACCGGCGACATCTGTTGGTCTTCGGAG 1514
Qy 1391 CAGTCCTTCGAAGCTTATGTTAAATGCCACGGTGAAA-----CCAA 1432
Db 1515 GAGATGAGATTCGTAGGTATATCTCTGAATCATCCAAATCTCGTTTCATCGAAACATCCAA 1574
Qy 1433 AGCGCGGATATTTTGGTGTGATCGGTGATTTGGAGGTCACGAGCACCGAGGTGGCTC 1492
Db 1575 CGGCTACGATGTTTTCAAAGGTACCCGACTTGGGATCCGACCTGACCGGTTGTTGTCAT 1634
Qy 1493 AGTTTTCAGCTCGAGGACCGAGTTTACGCAATCTTCGATACATAAAACCGGATGATG 1552
Db 1635 CTTCTCTGCTGCTGCTCTTAATCCAGAGCGCGGAGATCTTAAACCGGATGTAATCG 1694
Qy 1553 CTCGGGAGTCAATATCATTTGCGCTTGGCTCAAAATCTAGACCAACCGGACTTCCTT 1612
Db 1695 CACCTGGTTTGAATATTCTAGTCTGTTGGCTGACCGGATTTGTCATCTGGTGTACTT 1754
Qy 1613 ATGATTCGAAGAGTTAATCTTCACTGTATGTCAGGAATTTCAATGTCTGTGTCACATG 1672
Db 1755 CTGATAACCGGAGAATGAGTTCAACATTTTATCAGGCACTTCGATGGCTGCCGCGCAGC 1814
Qy 1673 TTAGCGGAATCACTGCTCTTATCCGCTGCTCATACCGGACTGCTCCAGCTGCAATCA 1732
Db 1815 TGTCTGCTAGTCTGCTTGTCTTAAGCGGCTCATCCGGATGGAGTCCGCGCGCGANTAA 1874
Qy 1733 AATCCGATTCATGACAAACCGGATTTGACGATCGTCAAGGAAGCAATAAAGGATG 1792
Db 1875 GATCAGCATTTGATTAACCAACCGCTTACAGGTTGATACAGCGGTGAGCCGATGATGATG 1934
Qy 1793 -----GTAACAAACCGCGGCTGTTTGGATTTGAGAGGCGCATGTGAATCCGCAA 1846
Db 1935 AGTCCACTGGCAATACATCTTCGGTTATGGATTATGTTCCGGTCACTGTTTCCACCAACCA 1994
Qy 1847 AGGCGATAACCCGGGATTTGTTTAAACATTAACAGGTTGATTAATCACTTACCTCT 1906
Db 1995 AAGCTATGGATTCGGGATTTAGTCTATGATATAACATCTTATGATTAATCACTTCTTGT 2054
Qy 1907 GCATCTTGGATTCACAGATTCAGATATTTTGAATCAATCACTTAAGAAAGCTGAGCTGCA 1966
Db 2055 GTAATCTTAATCACTAGAACCAACATTTGTGACAATAACCCGTGCGCAAGCGGACTGTG 2114

Qy 1967 ATGGAATATTCCGGAAAAACCCGGGTTTTAGTCTCAATTACCGTCTGATAGCCGTGATTT 2026
Db 2115 ACGTGTGAAGACGAGCGGACACGTCGGGAATTTAAACTACCGTCTGTTCTCTGTCGTGT 2174
Qy 2027 TCAA-----ACGTGGCAAGACTACGGAGATGATCAAGGCGGTGTCACTAACGTTG 2077
Db 2175 TTCAGCAGTAGGAGAGAGTAATAATGTCGACACATTTTATTAGGACTGTGACAAATGATG 2234
Qy 2078 GGAGTCCTTAATCGATATATCTCAGTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTG 2137
Db 2235 CGGATTCGGATTCGTTTACGAGATTAAAGATTAGCGCGCTAGAGGGACTACGGTGACGG 2294
Qy 2138 TCAATCTTAAGAGACTTGTGTTTCAACACGCTGGATCAGACGCTGAGCTATAGATATGTT 2197
Db 2295 TTGAGCCGAGAGACTATCGTTTCAAGCGGTGGGGCAGAAACTGAGTTTTTGTGTTAGGG 2354
Qy 2198 TT 2199
Db 2355 TT 2356
RESULT 10
AAC42012
ID AAC42012 standard; DNA; 2638 BP.
XX AC AAC42012;
XX AC AAC42012;
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 33960.
XX
DE Hybridisation assay; Genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 18-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152163.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match 17.9%; Score 417; DB 21; Length 2638;
Best Local Similarity 51.5%; Pred. No. 5.4e-119;
Matches.1148; Conservative 0; Mismatches 1030; Indels 53; Gaps 6;

Qy	13	CCTTCTCTTCTCTGCAATATCTTTCTCTCTATTTTGGTCTCTCTCTGTCAGAGATCTCTGCAG	72
Db	137	CCTAACTTTCTCTGCTCAATGGCTTTCTCTCTTACTTTTCTCTCTCTCTGACTCTTTCTTT	196
Qy	73	AAGCAGACTTACATTTGTTTCAGCTTCATCTCTAATAGCGAABACCGCTAABAACTTTGCCTCA	132
Db	197	CTCCTTCTCTCTGCTTCTCTCTTCAAACTCTTTGACTTATCATCGTCCACGTGACCCACG	256
Qy	133	AAGTTTGATTGGCATCTTTCTTTCTCCAAGAAGCGGTTTAGGTGTGTGAAGAAGAAGAG	192
Db	257	AAGCTAAACCTCAATCTTCCCCACTCATCTTCACTGGTACATTTCTCTCTTTCGCTCAC	316
Qy	193	GAAGAGC--CTTCTTCTCGACTTCTTACTCTCTATGGCTCTGCGAATTGAAGGATTTGCTG	250
Db	317	TCACATCTTCTCTCTCCCTCCATATACACACATACAACACGTTTTCCATGGCTTTCTCG	376
Qy	251	CTCAGTTGACTGAATTCAGAAGCGAGGATACGTGAGATATTCACCTGAAGTTGTTTGCAGTGA	310
Db	377	CTGACTCACTTTCACAAGACGTAGTCAACTCTTTAGACCATCTCTCATGTCACTCTCCGTTA	436
Qy	311	GACCTGACCATTGTTCTTCAGGTTCAAACCACTTACTCTTACAGTTCTTGGGACTCGACG	370
Db	437	TCCCGGAACAGTCCGTCACCTGTCACACCACTCGTTCCCTCGAGTTTCTTGGACTCAGGT	496
Qy	371	GTTTTGGAACTCCGGTCTATGGTCTAAATCTCGGTTTGGTTCAGGCGACAAATATCGGCG	430
Db	497	CCACCGCAAGACGGGCTTACTCGAAGAGTCTGATTTCCGGTCCGATCTAGTTATCCGGAG	556
Qy	431	TGCTTGATCTGAGGAGTTGGCTCGAAAGTCTCTAGCTTTGACGATACCGGAATGCTTCGA	490
Db	557	TTATCGATCTAGTGTGTTTGGCCCGAAGACCTAGCTTTGATGACCGTGGTCTTGGTCTG	616
Qy	491	TTCCACGGAAATGGAAGGGGATTTGCCAAGAAGAGAAAGTTTCAGTTCTTCGAGCTGTA	550
Db	617	TTCCCATTAATGGAAGGCCAATGTATCGCTTCCCAAGATTTTCCGGAGTCTGCTTGTA	676
Qy	551	ACCGGAAGCTAATCGGTGCTAGATCTTTCATCAGAGGACACCGTGTGCTTAATTCACAG	610
Db	677	ACCGTAAACTCGTCGGAGCTAGATTCTTCTCGCGTGGGTA-----TGAAGCAACCA	727
Qy	611	AGGAATCACAACACATGCCTCGTGAATACATTTCCGCAAGAGATTCACCGGACACCGGA	670
Db	728	ACGGGAAATGAATGAACACGACTGAGTTTCGCTCTCCGCGGTGACTCCGATGGACATGGGA	787
Qy	671	CTCACCGCGCTCAACAGTTGGTGGATCCTCTGTGTTTCGATGGCGAATGTTCTTGGCAATG	730
Db	788	CTCACACAGCTTCAATCTCCCGCGGCGGTACGTTTTCGGGCATCAACTCTCGGCTACG	847
Qy	731	GAGCTGGTGTGGCTCGTGGGATGGCTCTCGAGGCTCACTTGCAGTCTATAAAGTCTGT	790
Db	848	CTCAGGTTGCTGCTGCTGGGATGGCTCGAAGCTAGACTCGCTGCGGTACAAAGTCTGTT	907
Qy	791	GGTTCAATGGTTGTATACAGCTCTGACATCTTAGCAGCTATAGATGTAGCGATTCAAGATA	850
Db	908	GGAATCCGGTTGTTACGACTCAGATACTTTAGCGGCTTTTCGACACCGCGTTCGCGACG	967
Qy	851	AAGTCGATGTTCTTTTCGCTTTCCCTTTGGCGGTTTCCCTATTCTCTTTGTATGATGACACAA	910
Db	968	GTGTGATGTCACTCTCTCTCCGTTGGAGGCGTTGTGGTTTCCITTATACCTAGACGCTA	1027
Qy	911	TCGCCATTGGAAACATTCGAGGCCATGGAAACGGGTATATCTGTAACTGTGTCAGCTGGTA	970
Db	1028	TCGCCATTAGGAGCTTTTGGAGCTATTGACAGAGAAATTCGTCTCTGCTTCCGCGCGAA	1087

	Qy	971	ACAAAGCGTCAATCTGTAAGAAAGCTCTGTGGTCCAAACAACAGCTCCTTGTTGGTCTCAACCATCTGGCG	1037
	Db	1088	ACGAGGTCGCCGTGCCTTTAACGGTGACGAATGTGTCTCGTGATGACAAACAGTCCGAG	1147
	Qy	1031	CAGCACGCTTGATCGAAGATTTCCCGCTGTGFTCAGATTAGCCAACCGAAAAGCTTCTCT	1090
	Db	1148	CTGGAACAATCGATAGGGATTTTCCAGCCAATGTGAAACTCGGTAAACGGAAAGAATGATTT	1207
	Qy	1091	ATGAGAGAGTCATTGTATCCGGGAAAAAGGTATAAA-----GAATGCCGGGAGAGAGG	1141
	Db	1208	CTGGTGTTAGTGATACGGTGGACCGGGTCTGGATCCGGGTGGAATGTAACCGCTGTGTT	1267
	Qy	1142	TTGAGGTGATTTACGTCAAGGAGGAGATAAAGGAAGTGAATTCCTGTTTGAGAGGTCAC	1201
	Db	1268	ACGCTGTAGTCTTACTAGGCGGGGATGGTTACTCATCGTCTGTGTCTTTGAAGGCTCGT	1327
	Qy	1202	TTCCAAGAGAAGAAATCCGAGGCAAAATGTGTATTTGTGATCGCGGAGTCAAATGGAAGAT	1261
	Db	1328	TGGATCCGAATTTGGTGACCGGAAAGATCGTCTTTGTGATAGAGGAATCAATCTTAAG	1387
	Qy	1262	CGGAGAAGGAGAACCGGTTTAAAGAAAGCTGGAGAGTTGCAATGATCTTAGCCAATACAG	1321
	Db	1388	CAACCAAGGTGAGATCGTACGGNAAAATGGAGSCTTGGGATGATTTAGCGAATGGTG	1447
	Qy	1322	AGATCAACCAGAAGAGATCTTATTGACGTTTCATCTCTTACAGCTACATTTGATTCGTT	1391
	Db	1448	TGTTTGACGGTGAAGGTTTAGTAGTCTGATTTGCCACGTTTACCGGCACATCTGTTCGGTG	1507
	Qy	1382	ACACTGAGTCAGTCCITCTGAGGCTTATGTTAATGCCACGGTGAA-----	1428
	Db	1508	CTTCTGGAGAGATGAGATTCTGTAAGTATATCTCTGAATCATCCAAATCTGTTTCATCGA	1567
	Qy	1429	----CCAAAGGCGCGGATAATTTTTGGTGGTACGGTATTTGGAGGTACAGACACCGG	1483
	Db	1568	AACATCAACGGCTACGATTTGTTTTCAAGGGACTCGACTTGGATTCGACCTGCACCGG	1627
	Qy	1484	AGGTGGCTCAGTTTTTCAGCTCGAGGACCGAGTTTAGCCAATCCTTCGATACTAAACCGG	1543
	Db	1628	TTGTTGCATCTTTCTCTGCACGTGTCCTAATCCAGAGACGCGGAGATTCTTAAACCGG	1697
	Qy	1544	ATATGATGTCTCOGGGAGTCAATATCATTTGGGGCTTGGCCCTCAAAATCTAGGACCAACCG	1603
	Db	1688	ATGTAATCGCACCTGGTTTGAATATTTAGCTGCTTGGCCCTGACCGGATTGGTCCATCTG	1747
	Qy	1604	GACTTCTCTTATGATTCAGAGAGAGTTAACTTTCACTGTAATGTGAGGAACCTTCAATGCTT	1663
	Db	1748	GTGTACTCTGTATAACCGGAGAACTGAGTTCAACATTTTATCAGGCACCTTCGATGGCGT	1807
	Qy	1664	GTCCACATGTTAGCGGAATCACTGCTCTTATCCGGTCTGCATATCCCGAACTGGTCTCCAG	1723
	Db	1808	GCCCTCATGTTCTGGTCTAGCTGCTTTGCTTAAAGCGGCTCATCCGGAATTGGAGCCCGG	1867
	Qy	1724	CTGCAATCAATCCGATTTGATGACAAACAGCGGATTTGTACGATCGTCAAGGGAAACGGA	1783
	Db	1868	CGGCGATAAGATCAGCATTTGATGACCACGGCTTACACGGTTTGATAACAGCGGTCGAGCCGA	1927
	Qy	1784	TAAAGGATGG-----TAAACAACACCGCGGTGTGTTTGGATTTGGAGCAGGGCATGTGA	1837
	Db	1928	TGATGGATGAGTCACACTGGCAATACGTTCTCGGTTACGGAATATGGTTCCGGTCAATGTC	1987
	Qy	1838	ATCCGCAAAAGCGGATATAACCCGGGATGGTTTTACAACATTCACACAGTGGATTACATAA	1897
	Db	1988	ACCAACCAAGCTATGATCCGGGACTAGTCTACGATATAACATCTTATGATTATATCA	2047
	Qy	1898	CTTACCTCTGCACCTCTTGGATTCAAGAATCAGATATTTTAGCAATCACTCATAGAACG	1957
	Db	2048	ACTTCTTGTGTAATCTCTAACTPACACTGGAAACCAACATTTGTACGATAAACCGTCCGCAAG	2107
	Qy	1958	TGAGCTGCATTGGAAATATTTCGGGAAAAAACCGGGTTTTAGTCTCAATTAACCGTCSGATAG	2017
	Db	2108	CGGACTCGCAGCGGTGCGAGGCGAGCGGGAACAGTTCGGAAATTTGAACTACCCGTCGTCT	2167

QY 2018 CCGTGATTTTCAA-----ACGTGGCAAGACTACGAGATGATCACAAGCGGTGTCA 2068
Db 2168 CTGTCGTGTTTCAGCAGTATGAGAGAGTAAATGTCACACATTTTCATTAGGACTGTGA 2227
QY 2069 CTAAGCTGGGAGTCTTAACCTGATATCTACTAGTGAATGTCAAGCTCCAGAGGGATCA 2128
Db 2228 CAAATGTAGGCGATTTCGATTTCGATTTCAGAGATTAAAGATTAGGCGGCTAGAGGGACTA 2287
QY 2129 AAGTTATTGTCATCTTAAGAGACTTGTGTTCAAAACAGTGTGATCAGACGCTGAGCTATA 2188
Db 2288 CAGTGACGGTTGAGCGGAGAGACTATCGTTTCAGACGGGTGGGCGAAGAACTGAGTTTGTG 2347
QY 2189 GAGTATGGTTT 2199
Db 2348 TTGTTAGGTTT 2358

RESULT 11

AB213381
ID AB213381 standard; DNA; 2295 BP.

XX
AC AB213381;

XX
DT 21-JAN-2003 (first entry)

XX
XX Arabidopsis thaliana stress regulated gene SEQ ID NO 1186.

DE Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.

Matches 1072; Conservative 0; Mismatches 898; Indels 45; Gaps 9;
QY 205 TCTCGACTTCTCTACTCTCTATGGCTCTGGATTGAAGATTTCTGCTCAGTTGACTGAA 264
Db 166 TCTCGATCGTCCATGTTTACACACAGTCTTCCATGTTTCTCGCGCGTGTACTTCA 225
QY 265 TCAGAAGCGAGATATGAGATATTCACCTGAAGTTTGTGAGTGAAGACCTTACCATGTT 324
Db 226 GATGAAGCAGATAATCTCGTAAACCAACCCAGCAGTTCTTGTGTTTTCGAAGACCGA 285
QY 325 CTTTCAGGTTCAACCACTTACTCTTACAGTTTCTTGGGACTCGAGGTTTTCGAACACT 384
Db 286 CGAGAGCTTACACACACAGTCTCTCTCAATTTCTTGGTTTACAA-----AACCAAAA 339
QY 385 GGTGTATGGTCTAAATCTCGGTTTGGTCAAGGACAAATATCGGGTGTCTGATCTGGA 444
Db 340 GGAATATGGTCAAGTCTGATACGATCAGACGATATCATTTGGCGTTTTCGACACCGA 399
QY 445 GTTGGCTGAAAGTCTAGCTTTGAGATACCGGAATCCCTTCGATTCACCGGAATGG 504
Db 400 ATTTGGCGGAGCGGAGGAGTTTCTCAGATCTTAACCTCGTCCCAATTCCTCAAAAAG 459
QY 505 AAAGGATTTCCCAAGAGAGAAATTTTCACTTCTTCGAGCTGTAAACCGGAGCTAATC 564
Db 460 AGAGGCGTTTTCGAATCCGAGGCCAGATTCAGTCTCGGAACCTGTAAACCTAAATAT 519
QY 565 GGTGTAGATTTCTTATCAGAGGACACCTGTCGCTAATTCACAGAGAAATCACCACAA 624
Db 520 GGAGCAAGATTTCTCGTAAGGACAAACAGCGCT--GTAATCGGAGAAATCAACAAA- 576
QY 625 ATGCTCGTGAATACATTTCCGCAAGAGATTTCAACCGGACACGGGACTCACACCGCTCA 684
Db 577 ---ACCGTTGAGTTTCTATCTCTCTGAGCGGATGGACAGGTACTCACACTTCTCA 633
QY 685 ACAGTTGGTGGATCTCTGTTTCGATGGGGAATGTTCTTGGCAATGGAGCTGTGTGGCT 744
Db 634 ACCGCGCTGCGGCTCACGCTTTTAAAGCGAGTATGTCGCGTTTACGCTTCCGCTGTAG 693
QY 745 CTTGGGATGCTCTCGAGCTCACATTCGAGTCTATTAAGTCTGTGGTTTCAAT---GGT 801
Db 694 AAAGGTGTTGCTCCAAAGCTCGTATCGCGGCTTACAAAGTCTGTGGAAAGATTCGGT 753
QY 802 TGTACAGCTCTGACATTTAGCAGCTATAGATGTAGCGATTTCAAGATAAGTTCGATGTT 861
Db 754 TGCTCGATTCGATATTTCTCGCGCTTTGATGCGCGTGTAGAGACGCTGTGCGAGTT 813
QY 862 CTTTCGCTTTCCCTTGGCGTT-----TCCCTATTCTTTGATGATGACACATC 912
Db 814 ATATCGATCTCAATCGGTGCTGGAGACGGGATTTACTTGGCGTATTACCTCGATCCAAT 873
QY 913 GCGATTGGAATTCGAGCCATGGACGGGTATATCTGTAATCTGTGAGCTGTGTAAC 972
Db 874 GCTATAGGCTCGTACGCGCGCGCTCGAAAGGAATCTTCTCTCTCTCTGCGCGGAAC 933
QY 973 AACGTCCTCAATCGAAAGCTCTGTTGCAACACACAGCTCTTGGGTCTCAACCATTTGG 1032
Db 934 GAAGAGCTTAAGGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 993
QY 1033 GGCAGCTTGCATGAAGATTTCCGCTGTGCTGATGATGATGATGATGATGATGATGAT 1092
Db 994 AGTCAATCGATCGAATTTCCAGCGGATGCTATTCTCGCGACGAGACATCGTCTCAGA 1053
QY 1093 GGAGAGCTATGATTCGGGAAAGGATTAAGATTCGCGGAGAGAGGTTGAGTGTATT 1152
Db 1054 GGAGTGTCTCTTTACGCTGGAGTACCTTTAAACCGTCTGATGTTTCCGCT---GGTTAT 1110
QY 1153 TACGTACAGAGAGATTAAGGAGTGTGTTGTTGAGAGGTTCACTTCCAGAGAA 1212
Db 1111 CCCGTAATTCGGGAATGTCTATGATGATGATGATGATGATGATGATGATGATGATG 1170
QY 1213 GAAATCCGAGGCAAAATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 1272
Db 1171 CAAGTGGGGTAAATAGTAAATCTCGATAGAGAGAGCAGTCCACGCGTAGCCAAAGGA 1230

Query Match 17.1%; Score 398.2; DB 24; Length 2295;
Best Local Similarity 53.2%; Pred. No. 3.7e-113;

Qy	1273	GAACGGTTAAAGAAAGCTGGAGGAGTTGCAATGATCTTAGCCAAATACAGAGATCAACAA	1332
Db	1231	TTGCTTGTGAGAAAAGCAGGTGGTGTCCGAATGATCTCGCTAATGAGCATCTTAACGGT	1290
Qy	1333	GAAGAAGATCTATTGACGTTCACTCTTACCACTACATGATGTTGTTACACTGATCA	1392
Db	1291	GAAGATTATGCGGAGATGCTCATCTATTCCAGCTGTCGGTTGGATCAAAAGGAAGGA	1350
Qy	1393	GTCTCTCTGAAGGCTTATGTAAATGCCACGCTGAACCAAGGCGCGGATAATTTTGGT	1452
Db	1351	GATAGATCAAAGCATATGCTTCTTCACATCCGATCCAAATGCTTCAATTGATTCAGA	1410
Qy	1453	GGTACGCTGATTTGGAGGTCACGAGCACCGAGGTGGCTCAGTTTTCAGTCTCGAGGACCG	1512
Db	1411	GGAACATAGTTGGGATTAACCGGCTCCGGTTATTGCTTCTCCGTTAGAGGACCA	1470
Qy	1513	AGTTTAGCCAATCCTTCGATACATAAACCAGGATGATTCGTCGGGAGTCAATATCAT	1572
Db	1471	AACGGTTAAGCCCGAGATCTTTAAACCGGATTTGATTCGCCGAGTTAAATCCTC	1530
Qy	1573	CGCGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTCAGAAAGATTAAC	1632
Db	1531	GCCGATGGACAGACGCTGTGGACCTACAGTTTGCGTCAGATCCAGGAATAACGGA	1590
Qy	1633	TTCACTGTAATGTCCAGGAATCTCAATGCTTTGTCCACATGTTAGCGGAATCACTGCTCT	1692
Db	1591	TTCAACATCTCTCCGTFACCTTCAATGGCATGCTCAGCTAGTGTGGCGCGCGCTT	1650
Qy	1693	ATCCGCTTCGATACCCGAAGTGTGCTCCAGCTCAATCAAAATCCGCAATGATGACAACA	1752
Db	1651	CTCAATCCGCTCATCCAGATTTGAGCGCTCCCGTGATAGATCGGCAATGATGACAACG	1710
Qy	1753	CGCGATTGTACGATCGTCAAGGAAAGCGATAAAGGATGGTAAC-----AAACAGCC	1806
Db	1711	ACTAACTCGTCGATAACTCTPAAACCGCTCGTTGATCGATGAATCCACCGGGAATCGGCT	1770
Qy	1807	GGTGTGTTTGGCATTTGAGCAGGCGCATGTGAATCCGCAAAAGGCGATAAACCCGGGATTG	1866
Db	1771	ACGCGTTATGATTAACGGGTGGGTCACTTAATTTGGCGCGGCTATGAATCCGGGTCTT	1830
Qy	1867	GTTTACAACATTC AACCAGTGGATTACATAACTTACCTCTGCATCTTTGGATTCAACA	1926
Db	1831	GTCTACGATATAACTAATGATGATTAATTAAGTTCTTTGCTCCATCGGTACGGACCA	1890
Qy	1927	TCAGATATTTAGCAATCACTCATAGAAAGTGAGCTGCAATGGAATATTGCGGAATAAC	1986
Db	1891	AAGACGATCCAAGTGATAAAGAAACACCGGTGAGATG--TCCGACGACCGAGGAACCG	1947
Qy	1987	CCGGTTTTTAGTCTCAATTAACCGTCGATAGCCGTGATTTTC-----AAACGTGGC	2037
Db	1948	TCTCCCGGGAATTTGAATATCTTTCGATCAGCGCGGTGTTTCTACTAATAAGAGAGGA	2007
Qy	2038	AAGACTACGGAGATGATCAAGGCGTGTCACTAACGTTGGAGTCTCTAACTCGATATAC	2097
Db	2008	TTGCTGATTAACCTGTTTATAGGACCGGACGATGTCGGGACGGCTGAGCGGTTTAT	2067
Qy	2098	TCAGTGAATGTCAAGGCTCCAGAGGGATCAAAAGTTATTGTCAATCTTAAGAGATTGTG	2157
Db	2068	CGGCGGAGGATAGAGTCGCGAGAGGAGTGACGCTGACAGTGAACCACTAGGCTTTGTG	2127
Qy	2158	TTCAACAACGTTGGATCAGACGCTGAGCTATAGAGT	2192
Db	2128	TTTACTTTCCGCGGTTAAGAGACGGAGCTATGCGGT	2162

RESULT 12

RESULI 12
AAC42230

AAC42230
 ID AAC42230 standard; DNA; 2489 BP.

XX

AC AAC42230;

XX

DT 17-OCT-2000 (first entry)

XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 34769.
DE	Hybridisation assay; genetic mapping; gene expression control;
XX	protein identification; signal transduction pathway;
KW	metabolic pathway; promoter; termination sequence; ss.
KW	
XX	
OS	Arabidopsis thaliana.
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0131825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0137462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131443.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	25-MAY-1999; 99US-0136382.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139460.
PR	18-JUN-1999; 99US-0139461.
PR	18-JUN-1999; 99US-0139462.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	18-JUN-1999; 99US-0139763.

10

QY 505 AAAGGATTTGCAAGAGGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAGCTAATC 564
Db 654 AGAGGGTTTGGCAATCCGAGGACAGATTCACTTCCTCGGAACTGTAACCGTAAATATC 713
QY 565 GGTGCTAGATTCTTCATCAGAGACACCGTGTGCTAATTCACACAGAGAAATCACCAAC 624
Db 714 GGAGCAAGATTCTTCGCTAAGGACAAACAGCGCT- -GTAATCGAGGAATCAACAA- 770
QY 625 ATCCCTCGTGAATACATATTCGCAAGAGATTCAACGGGACACGGGACTCACACCGCCTCA 684
Db 771 - - -ACGTTGAGTTTCTAATCTCTGTCGCGGATGGAACCGGATGGAACCTTCCTCA 827
QY 685 ACAGTTGGTGGATTCCTCTGTTTCGATGGGAATGTTCTTGGAATGGAGCTGGTGGCT 744
Db 828 ACCGCCGTGGCGCTCAAGCTTTTAAAGCGAGTAAGTCCGTTACGCTTCGCTGAGCC 887
QY 745 CGTGGAGTGGCTCTCGAGCTCAATTCGAGTCTAATAAGTCTGTGGTTCAAT- -GGT 801
Db 888 AAAGGTGTTGCTCCAAAGCTCGTATCGCGGCTACAAAGTCTGTTGGAAGATTCCGCT 947
QY 802 TGTTACAGCTCTGACATTTAGCAGCTATAGATGTAGCGATTCAAGATAAAGTCGATGTT 861
Db 948 TGTTCTGATTCCGATATTTCTGCGGCTTTGATGCGCTGTGTAGACAGGTTGTGACGTT 1007
QY 862 CTTTCTGCTTTCCCTTGCGGCTT- - - - -TCCCTATTCTTTGTTATGATGACACAATC 912
Db 1008 ATATCGATCTCAATCGTGTGGAGACGGGATTACTTCGCGTATTACCTCGATCCAATC 1067
QY 913 GCCATTGGAAATTCGAGCAGTGAACCGGCTATATCTGTAATCTGTCAGCTGTAAC 972
Db 1068 GCTATAGGCTCGTACGGGCGGCTCGAAAGGAATCTTCGTCTCTTCCTCTCGCGGAAAC 1127
QY 973 AACGGTCCAATCGAAAGCTCTGTTGCAACACAGCTCTTGGTCTCAACCAATTTGGCA 1032
Db 1128 GAAGGACCTAACCGTATGTCACTACGAACTCGCGCGTGGGTAAACCACTGCTGCT 1187
QY 1033 GGACGCTGATCGAAGATTTCCGCTGTGCTGATAGCAATAGCAACGGAAGCTTCTAT 1092
Db 1188 AGTACAATCGATCGGAATTTCCAGCGGATGCTATTCTCGGCGACGGAATCGTCTCAGA 1247
QY 1093 GGAGAGTCATTGTTATCCGGGAAAGGTATAAAGATTCGCGGAGAGAGGTTCAGGTGATT 1152
Db 1248 GGAAGTCTCTTACGCTGGAGTACCTTTAAACGGTCTGATGTTCCCGT- - -GGTTAT 1304
QY 1153 TAGCTCAGAGGAGATAAAGAGTGTGATTCGTTGAGAGGTCATTCACAGAGAA 1212
Db 1305 CCCGGTAAATCGGGAATGTATCAGCGTCTCTATGTATGGAGAACACACGCTTCATCCGAAG 1364
QY 1213 GAATCCGAGGCAAAATGGTATTTGTATCGCGGAGTCAATGGAAGATCGGAGAAAGGA 1272
Db 1365 CAAGTGGGGTAAATAGTAATCTCGATAGAGGAAGCAGTCCACGCTAGCCAAAGGA 1424
QY 1273 GAAGCGTTAAAGAGCTGAGGAGTTGCAATGATCTTAGCCAAATACAGAGATCAACCA 1332
Db 1425 TTGGTTGTGAAGAACAGAGTGTGTGCGAATGATTTCTGTAATGGAGCATCAACGCT 1484
QY 1333 GAAGAGATTTCTATTGACGTTCTATCTTACCAGCTTACATTTGATTTGGTTACACTGATCA 1392
Db 1485 GAAGGATTAGTCGAGATGCTCATCTATTCCAGCCTGTGCGCTTGGATCAACAGAGGA 1544
QY 1393 GTCCTCTGAAGGCTTATGTTAATGCCAGCGTGAACCAAGCGCGGAGTAATTTTGGT 1452
Db 1545 GATAGAATCAAGCATATGTTCTTCAATCCGAATCCAAATGCTTCAATTTGATTTTCA 1604
QY 1453 GGTACGGTATTCGGAGTTCAGGACCGCGAGTGGCTCAGTTTTCAGCTCGAGGACG 1512
Db 1605 GGAACATAGTTGGGATTAACCGGCTCCGGTTATTGCTTCTTCTCCGGTAGAGACCA 1664
QY 1513 AGTTTACCAATCTTCTGATCTATAAACCGGATATGATTTGCTCCGGAGTCAATATCAT 1572
Db 1665 AACGGTTAAGCCCGGAGATCTTAACCGGATTTGATTTGCTCCCGAGTTAATATCTC 1724

QY 1573 CGCGCTTGGCTCAAAATCTAGGACCAACCGGACTTCCATTATGATTCAGAAGAGTTAAC 1632
Db 1725 GCCCATGAGCAGACGCTGTGTGACCTTACAGGTTTCCGCTCAGATCAAGGAAACCGAA 1784
QY 1633 TTCACTCTAATGTTCAGGAATCTCAATGCTTGTGTCACATGTTAGCGGAATCACTGTCTTT 1692
Db 1785 TTCAACATTTCTCCGCTACTTCAATGGCATGCTTCACTGATTTAGTGGTGGCGGCTT 1844
QY 1693 ATCCGCTCTCATACCCGAATCTGCTCTCAGCTGCAATCAAAATCCGCAATGATGACAACA 1752
Db 1845 CTCAATCCGCTCATCCAGATTGGAGCCCTGCGTGATAGATCGGCAATGATGACAACG 1904
QY 1753 GCGGATTTGATCGATCTCAAGGAAACGATAAAGATGTTAAC- - - - -AAACGAGCC 1806
Db 1905 ACTAACCTCTGATGATCTTAACCGCTCTGATCGATGATGATGATGATGATGATGATGAT 1964
QY 1807 GGTGCTTTCGATTTGAGCAGGCGCATGCAATCCGCAAGGCAATAAACCGGATTTG 1866
Db 1965 ACGCTTTATGATTTACGGCTCGGCTCAATTAATTTGGCGCGGCTATGAATCCGGGCTT 2024
QY 1867 GTTTACAACATTCACCCAGTGGATTAATAACTTACCTCTGCACCTCTTGGATTCAAGA 1926
Db 2025 GTCTACGATATACTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2084
QY 1927 TCAGATATTTAGCAATCACTCAATAGAACGTCGATGCAATGCAATGCAATGCAATGCA 1986
Db 2085 AAGACGATCCAAAGTGAATCAAGAACACCGGTGAGATG- - -TCCGACGACGAGAAACCG 2141
QY 1987 CCGGCTTTTGTCTCAATTAACCGTTCGATAGCGGTGATTTTC- - - - -AAACGTGGC 2037
Db 2142 TCTCCCGGGAATTTGAATTTATCTCTGATCAGCGCGGTGTTTCTCTACTAATAGAGAGGA 2201
QY 2038 AAGACTACGAGATGATCAACAGCGGTGTCATCAAGCTTGGAGTCTTAACCTCGATATAC 2097
Db 2202 TTGTGATGATAAATCTGTTATAAGAACGCGACGAAATGTCGGGAGGCTGAGCGGTTAT 2261
QY 2098 TCAGTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTGTCATCTTAAGAGACTTTGT 2157
Db 2262 CGGCGGAGGATAGATGCGCGAGGAGTACGCTGACAGTGAACACCTAGGCTTTGTG 2321
QY 2158 TTCAAAACGCTGGATCAGACGCTGAGCTATAGAGT 2192
Db 2322 TTTACTTCCGCGCTTAAGAGACGAGCTATGCGGT 2356

RESULT 13

AAC32938
ID AAC32938 standard; DNA; 1653 BP.
XX
AC AAC32938;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 1197.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 26-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140355.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 08-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

```
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160788.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.2%; Score 285; DB 21; Length 1653;
Best Local Similarity 53.1%; Pred. No. 7.7e-78;
Matches 684; Conservative 0; Mismatches 595; Indels 18; Gaps 3;

QY 962 CAGCTGTAAACAAGCGTCCAAATCGAAAGCTCTGTGTGCAACACAGCTCCCTTGGTCTCAA 1021
DB 2 CTGCTGTAAATAGTGGCTCTACTAGAGCTTCTGTGTGTAATGTTCCTTGGTGTATGA 61

QY 1022 CCATTGGCGGACGCGCTTCATCGAAGATTTCCGCTGTGTGTCAGATTAGCCAAACGGAA 1081
DB 62 CTGTTGGTGGTACTTTAGATAGAGATTTTCGGCTTTTGGCAATCTCGTAAACGGGA 121

QY 1082 AGCTTCTCTATGGAGAGTCATTGTATCCGGAAAAGGTATAAGAAATGCCGGGAGAGAGG 1141
DB 122 AACGACTTACCGGTGTTTCGTGTATAGCGGTGTGGGAATGGGA-----CGAAGCCGC 175

QY 1142 TTGAGGTGATTTACGTACAGAGAGAGATAAAGGAAGTGTCTGTGTGAGAGGTCAC 1201
DB 176 TTGAATTGGTTTATAATAAAGGAATAGTAGTTCGAGTAATCTTTGTTTACCTGGTTCGC 235

QY 1202 TTCAGAGAGAGAAATCCGAGGCAAAATGGTGTATTCGTCGCGGAGTCAATGGAGAT 1261
DB 236 TTGATTCGAGTATTGTTCTGCGGAGAGATTGTTGTTGTGATAGAGGTGTTAATGCTAGG 295

QY 1262 CGGAGAAAGGAGAGCGGTATAAGAAAGCTGAGAGAGTTGCAATCATCTTAGCCAAATACAG 1321
DB 296 TTGAGAAAGGAGCTGTGTTTAGATGCTGTGTGTTTAGGATGATTAATGGCGAATACG 355

QY 1322 AGATCAACAAGAGAGAGATCTAATGACGTTTCATCTCTTACCAGCTACATTTGTTGTT 1381
DB 356 CTGCGAGTGGAGAGAGCTTGTGGCGGATGATCAATTCCTCCCGCATGCTGTAGGGA 415

QY 1382 ACACGTAGTCACTCTCTGAGAGGCTTAATGTAATGCAACGCGTAAACCAAAGCGCGGA 1441
DB 416 AGAAGACTGGTGATTTACTTAGGAGATGTTAAGTCAGATTCTAATCCAACTGCTCTTC 475

QY 1442 TAAATTTTGGTGTACGCTGATTTGGAGGTACAGAGACCGGAGTGGCTCAGTTTTCAG 1501
DB 476 TTGTTTTTAAAGGAACGGTCTTTGACGTTAAGCCGCTCTCTGTGTGCTGCTTTTACGT 535

QY 1502 CTCGAGGACCGAGTTTAGCCATCTCTCGATACCTTAAACCCGAGATGATTTCTCCGGGAG 1561
DB 536 CGAGAGGTCCTAATACTGTTACTCTCTGAACTTGAAGCCGTGATGTTATGTTGCTCTGAG 595
```

```
QY 1562 TCAATATCATTTGGCGCTTGGCCTCAAATCTAGGACCAACCGGACTTCCCTTATGATCAA 1621
DB 596 TGAACATTTTGGCTGGTGGTCTGACGCTATTGGTCTCTACTGGTCTTGACAGGACTCTA 655

QY 1622 GAAGAGTTAACTTCACTGTAATGTGAGAACTTCAATGTCTTGTCCACATCTTTAGCGGAA 1681
DB 656 GGAGGACTCAGTTTAAACATCATGTGAGTAGCTCAATGTCAATGCCACACATCAGTGGTT 715

QY 1682 TCATGCTCTTATCCGCTCTGCAATACCGGAATCGGTCTCCAGCTGCAATCAAAATCCGAT 1741
DB 716 TAGCGGTCTTTTGAAGCAGCTCACCTGAGTGGAGTCCGAGTGTCTATCAATCAATCAGCTC 775

QY 1742 TGATGACAAACAGCGGATTTGTACGATCGTCAAGGGAAGCGATAAAGATCGT-----A 1795
DB 776 TCATGACTACAGCTTACGTTCTTGACAAACCAACGCTCCTCTCATGATGCTGCAGACA 835

QY 1796 ACAAAACCAAGCGGTGTGTTGCGATTGGAGCAGGCGCATGTAATCCGCAAAAGCGCATAA 1855
DB 836 ACAGCCTATCTTAACCATATGCTCATGGCTCGGCGCATGTAGATCCCCAAAGAGCTCTCT 895

QY 1856 ACCCGGGATTTGGTTTACAAACATTTCAACAGTGGATTACATAACTTTACCTCTGCACTTTG 1915
DB 896 CACCAGGTCTTGTCTACGACATCTCAACCGAGGAATACATCAGGTTTTTGTGCTCTCTAG 955

QY 1916 GATTCACAAGATCAGATATTTTAGCAATCACTCATPAGAACGTGAGCTGCAATGGAATAT 1975
DB 956 ACTACACAGTCGATCACATTTGTTGGAT-----TGTGAAGCGACCTAGCGTTAACTGCT 1009

QY 1976 TCGCGAAAAACCCGGGTTTTAGTCTCAATTTACCGTCCGATAGCCGTGATTTTCAAACGTG 2035
DB 1010 CGAAGAAAGTTCTCAGATCCTGCTCAGCTCAACTACCCAAAGTTTTCGGTTTTTGTGGGG 1069

QY 2036 GCAAGACTACGGAGATGATCAAGGCGGTGTCACCTAACGTTGGGAGTCTTAACTCGATAT 2095
DB 1070 GTAAAGAGTGTGCGGTACACTCGGGAAGTAAACAAATGTTGGTGCAACAGCTCGGTTT 1129

QY 2096 ACTCAGTGAATGCAAGGCTCCAGAGGGGATCAAAAGTTATTGTCAATCTTAAGAGACTTG 2155
DB 1130 ACAAGTGACGTTTAAATGGAGCTCTAGTGTGGAATCTCTGTTAAACCATCGAAACTTT 1189

QY 2156 TGTTCAAACACGTGGATCAGCGCTGAGCTATAGAGTATGTTTGTATTTAGAGAGAAA 2215
DB 1190 CGTTTAAAGCGTGGGAGAGAGAGGTACACAGTCACGTTTGTAGCAAGAAAGAG 1249

QY 2215 ACAGAGGAGGAGGTGGCTAGCTTTG 2242
DB 1250 TGAGTATGACGAACAAGGCTGAGTTG 1276

RESULT 14
AAC42099
ID AAC42099 standard; DNA; 2402 BP.
XX
XX AAC42099;
XX AC
XX AAC42099;
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 34284.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX BP1033405-A2.
XX
XX 05-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
```

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135829.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 13-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154773.
PR 22-SEP-1999; 99US-0155133.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.


```
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 09-OCT-1999; 99US-0158369.
PR 10-OCT-1999; 99US-0158369.
PR 11-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158369.
PR 14-OCT-1999; 99US-0158369.
PR 15-OCT-1999; 99US-0158369.
PR 16-OCT-1999; 99US-0158369.
PR 17-OCT-1999; 99US-0158369.
PR 18-OCT-1999; 99US-0158369.
PR 19-OCT-1999; 99US-0158369.
PR 20-OCT-1999; 99US-0158369.
PR 21-OCT-1999; 99US-0158369.
PR 22-OCT-1999; 99US-0158369.
PR 23-OCT-1999; 99US-0158369.
PR 24-OCT-1999; 99US-0158369.
PR 25-OCT-1999; 99US-0158369.
PR 26-OCT-1999; 99US-0158369.
PR 27-OCT-1999; 99US-0158369.
PR 28-OCT-1999; 99US-0158369.
PR 29-OCT-1999; 99US-0158369.

Query Match 9.8%; Score 228.2; DB 21; Length 2402;
Best Local Similarity 48.3%; Pred. No. 5.6e-60;
Matches 977; Conservative 0; Mismatches 963; Indels 81; Gaps 9;

QY 211 CTTCTCTACTCTGCTCTGCGATTGAAGGATTTCTCTCAGTGTGCTGATCAGAA 270
DB 237 CTTCTCAGCATATAGCATGCTCTCAGGTTTTCAGGTTTTCAGTGTGCTGATCAGAA 296
QY 271 GCCAGATACCTGATATACCTGGAAGTTTGTGAGTGAAGCTGAGCTGACCATGTTCTTCAG 330
DB 297 GCCAGTCTATAGCAAGAACCGGAGTGTGTTTTCAGTGTGCTGATCAGAACTTCCAG 356
QY 331 GTTCAAAACCACTTACTCTTCAAGTCTTGGGACTCGACGGTTTGGAACTCCGGTGA 390
DB 357 CTTCTACAACTCATTCATGGGACTTCTCAAGTACCAACATCAGTAAAGGTCGATTC 416
QY 391 TGGTCTAAATCTCGGTTTGGTCAAGGACA-----ATTATCGCGTGTGTGATACT 441
DB 417 GGTCCAGCTTCATCAGGCTCAGATGGAKCATATGATAGCATTTGCGGAATTTTGACACA 476
QY 442 GGAGTTGGCTGAAAGTCTTAGCTTTGACGATACCGGAATGCTTCGATTCCACCGGAA 501
DB 477 GGGATATGGCCAGAGTCAGAGGTTTCAATGACAAAGACATGGGTCCAATTCGCTCGG 536
QY 502 TGGAAAGGATTTGCCAAGAGGAGAAAGTTTTCAGTTTCTTCGAGCTGTAAACCGGAAGCTA 561
DB 537 TGGAAAGGATCATGATGGAAGCAAGGACTTCAAGTCTTCCAACTGTACAGAAAGATC 596
QY 562 ATCGGTGCTAGATTTCTTCATCAGAGGACACGGTGTGCTAATTCACAGAGGAATACCA 621
DB 597 ATTGGAGCAAGATACTACA-----AAAATCTCT 623
QY 622 AACATGCTCTGATATATTTCCGCAAGAGATTCACGGGACACGGGACTCACACCGGC 681
DB 624 GATGATGATTCAGAAATATATACCAAGGAGATCTCATCGGTTCAGGTTCTCATGTGTC 683
QY 682 TCAACAGTGTGTGATCTCTGTTTCGATGCGGAATGTTCTTGGCAATGAGGTGTTG 741
DB 684 TCCACCATAGCTGGATCTGCCGTGGAGATGCTCTTCTACTATGTTAGTTCGGGACT 743
```

```
QY 742 GCTCGTGGATGGCTCTCTGAGCTCATTGCGAGTCTATAAAGTCTGTTGGTTCATGGT 801
DB 744 CCAAAGGGAGGTTTCAAAAACGCTAGAAATCGCTATGTACAAAGTATCAATCCAGGGGA 803
QY 802 TGTACAGCTCTGACATTCAGCAGCTATAGATAGCATTCAAGATAAAGTCGATGTT 861
DB 804 TGCATGGCTCTCTATCTCTAGCTGCTTTCGATGATGCAATCCGAGATGGATGTT 863
QY 862 CTTTCGCTTTCCTTGGCGGTTTCCCTATTCT-----TTGTATGATGACACAATC 912
DB 864 CTAATCTCTCTCTGAGCTCCAGCATACGCTCGCATCGACTTGAACACTGATCTATT 923
QY 913 GCCATTGGAACATTCGAGGATCGAAGCGGATATCTGTATCTGTGAGCTGTGAGTAAAC 972
DB 924 GCCATTGAGCGTTTTCAGCGGTTGAGCAAGAAATCTTGTGTGATCTGCTCTCGCGGTAAT 983
QY 973 AACGGTCCCAATCGAAAGCTCTGTTGCAAAACACACTCTCTGGGCTCTCAACCATTTGGCGCA 1032
DB 984 GATGGACCTGATGCGGTTACTAATACTGCACTTGGATAATGACCCGTTGCTGCC 1043
QY 1033 GGCAGCTTGTATCGAAGATTTCCCGCTGTGTGATAGCATTAGCCAAAGGAAAGCTTCTCTAT 1092
DB 1044 AACACTATTGATAGACTTTTGTGATGTTTGTACTAGCGGCAATAAGTATCATCAAG 1103
QY 1093 GGAGAGTCAATGTATCCCGGAAAAGGTATAAA-----GAATGCCGGGAGAGAGGTT 1143
DB 1104 GGTGAAGGTAATACACTTTTCAACGTTAGTAAATCTCTGTGTATCTCTGATTCATGGC 1163
QY 1144 GAGTGTATTTAGTCAAGGAGGATTAAGGAAGTG----AGTCTGTTGAGAGGTTCA 1200
DB 1164 AAGTCTGCTAAGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1223
QY 1201 CTTCAAGAGAGAAATCCGAGGCAAAATGTTGTTGATCGCGAGTCAATGGGAAGA 1260
DB 1224 CTAGATCAAGAGAGGATTAAGGAAGATTTGTTATGCGAGAAAGTGTGTTGATCATAT 1283
QY 1261 TCGAGAAAGAGAGAGCGGTTTAAAGAAAGCTGGAGAGTTGCAATGATCTTAGCCAAATACA 1320
DB 1284 TATGATCATCTCGCTAGGACAAAGTGAAGAGCAAGAGGATCTGTTGGTCTTTGTA 1343
QY 1321 GAGATCAACCAAGAAAGATTTCTTATGACCTTCACTCTTTACAGCTTACATTTGTT 1380
DB 1344 GATGACAGAACTAGAGCAGTTGCTATGCTTATGCTTATGCTTATGCTTATGCTTATGCT 1403
QY 1381 TACACTGATCAGTCTCTTCAAGGCTTATGTTTATGCCAGGTTGAACCAAGAGCGCGG 1440
DB 1404 TCAAGAGAGAGCTGATGATCTTCTTACCTCACTCACTCACTCACTCACTCACTCACTCA 1463
QY 1441 ATAAATTTTGTGTTGCTGATGTTGGGAGGTCAGAGCACCGGAGGTTGCTCAGTTTCA 1500
DB 1464 ATCTTCCACTGCAAGTTGAAAGTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1523
QY 1501 GCTCAGGAGCGAGTTTACCCAATCTTCGATATCTAAACCGGATATGATTTGCTCCGGGA 1560
DB 1524 TCCAGAGACCTTCAAGCCTCACAAGAGCATTTCTCAAACTGACATTTACCGCACCGAGA 1583
QY 1561 GTCAATATATCTGCGGCTTGGCTCAAAATCTAGAACCAACCGGACTTCTTATGATTTCA 1620
DB 1584 GTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1640
QY 1621 AGAAGAGTTAACTTCACTGCTGATGTCAGAACTTCAATGCTTGTGTCACATGTTAGCGGA 1680
DB 1641 CCGGTTCTCTGATATACGTCATATCAGAACTTCCATGGAGCTCTCTCTCTCTCTCTCTCTCT 1700
QY 1681 ATCACTGCTCTTATCCGCTCTGCTATCCCGGAATCTGCTCCAGTGTGATCAATCAATCCOCA 1740
DB 1701 GTTGTATCTCTGATCAAAATCAGAGCTCCACATGGGGTCCATCCGCGATCAGATCAGCA 1760
QY 1741 TTGATGACAAACAGG---GATTTGATGATCGTCAAGGAAAGGATTAAGATGTTGAT 1797
DB 1761 ATTATGACAAACAGGACTTCAAAACAAACGACAAAGGTTTATAACCAACGAAACTGGT 1820
QY 1798 AAACACCGCGGTGTTGTCGATTTGGAGCAGGAGTGTGAATCCGCAAAAGCGGATTAAC 1857
```


us-09-806-767-1.rng

Tue Feb 3 10:49:41 2004

Job time : 460 secs

GenCore version 5.1.6
Copy-right (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 21:38:00 ; Search time 119 Seconds
(without alignments)
8634.789 Million cell updates/sec

Title: US-09-806-767-1
Perfect score: 2328
Sequence: 1 atggacacaaacctttttt.....taaccttgagactaactga 2328

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	62	2.7	7218	1 US-08-232-463-14	Sequence 14, Appl
2	58	2.5	293	4 US-09-313-294A-6881	Sequence 6881, Ap
3	44.8	1.9	1236	4 US-09-445-472-2	Sequence 2, Appl
4	44.8	1.9	1566	3 US-08-894-818B-4	Sequence 4, Appl
5	44.8	1.9	1962	3 US-08-894-818B-34	Sequence 34, Appl
6	44.8	1.9	1962	4 US-09-445-472-15	Sequence 15, Appl
7	44.4	1.9	295	4 US-09-313-294A-7407	Sequence 7407, Ap
8	37.2	1.6	289	3 US-09-007-005-17	Sequence 17, Appl
9	37.2	1.6	289	3 US-09-244-796-17	Sequence 17, Appl
10	37.2	1.6	3489	2 US-08-728-223A-1	Sequence 1, Appl
11	37.2	1.6	3489	4 US-09-298-568-1	Sequence 1, Appl
12	37.2	1.6	3489	4 US-09-410-399-1	Sequence 1, Appl
C 13	37.2	1.6	32207	2 US-08-770-379-20	Sequence 20, Appl
C 14	37.2	1.6	32207	2 US-08-757-669A-20	Sequence 20, Appl
C 15	37.2	1.6	32207	4 US-09-230-371A-20	Sequence 20, Appl
C 16	37	1.6	1230025	4 US-09-198-452A-1	Sequence 1, Appl
17	36	1.5	1859	3 US-08-894-818B-15	Sequence 15, Appl
18	36	1.5	1977	3 US-08-894-818B-2	Sequence 2, Appl
19	36	1.5	1977	3 US-08-894-818B-6	Sequence 6, Appl
20	36	1.5	1977	4 US-09-445-472-11	Sequence 11, Appl
21	36	1.5	2595	4 US-09-281-766-18	Sequence 18, Appl
22	36	1.5	7218	1 US-08-232-463-14	Sequence 14, Appl
C 23	34.6	1.5	938	4 US-09-205-258-211	Sequence 211, App
24	34.4	1.5	1146	2 US-08-504-365B-74	Sequence 74, Appl
25	34.4	1.5	1494	3 US-09-255-502-1	Sequence 1, Appl
26	34.4	1.5	1496	6 RE34606-1	Patent No. RE34,60
27	34.4	1.5	1497	1 US-08-322-677A-6	Sequence 6, Appl

28	34.4	1.5	1497	1 US-08-322-676-6	Sequence 6, Appl
29	34.4	1.5	1497	3 US-08-898-218-6	Sequence 6, Appl
30	34.4	1.5	1497	3 US-08-848-793-6	Sequence 6, Appl
31	34.4	1.5	1497	4 US-09-445-270-1	Sequence 1, Appl
32	34.4	1.5	1497	4 US-09-178-173A-1	Sequence 1, Appl
33	34.4	1.5	1497	4 US-08-322-678-6	Sequence 6, Appl
34	34.4	1.5	1497	6 5472855-1	Patent No. 5472855
35	34.4	1.5	1513	4 US-09-178-155-1	Sequence 1, Appl
36	34.4	1.5	1868	1 US-08-069-863-1	Sequence 1, Appl
37	34.4	1.5	1868	1 US-08-309-069-1	Sequence 1, Appl
38	34.4	1.5	1868	1 US-08-562-833-1	Sequence 1, Appl
39	34.4	1.5	1868	4 US-09-659-749-1	Sequence 1, Appl
40	34.4	1.5	1868	4 US-09-672-105-1	Sequence 1, Appl
41	34.4	1.5	1868	5 PCT-US95-05520-1	Sequence 1, Appl
42	34.4	1.5	8119	1 US-08-460-343B-1	Sequence 1, Appl
43	34.4	1.5	8119	1 US-08-398-028B-1	Sequence 1, Appl
44	34.4	1.5	8119	2 US-08-504-265B-1	Sequence 1, Appl
45	34.2	1.5	1867	1 US-07-772-087-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14
Query Match 2.7%; Score 62; DB 1; Length 7218;

	Query Match	2.5%;	Score 58;	DB 4;	Length 293;
	Best Local Similarity	54.3%;	Pred. No. 7.8e-09;		
	Matches 140;	Conservative 0;	Mismatches 115;	Indels 3;	Gaps 1;
QY	622	AACATGCTCTGTGAATACATTTCCGCAAGAGATCAACGGGACACGGGACTCACACCCCC	681		
Db	5	AACGCATCTCAGGATCTTGTCTTCCACTCTGATAGTGATGGCCATGGGACGATACACGA	64		
QY	682	TCAACAGTTGGTGGATCCTCTGTTTCGATCGCAATGTTCTTGGCAATGGAGCTCGGTGTG	741		
Db	65	TCAATTGCTGTGGCAATCATGGAATTCCTGTTGTTGTGCTGGACATCAATTTGGAAAT	124		
QY	742	GCTCGTGGATGGCTCTCTGGAGCTCACATTTGCAGTCTTATAAGTCTGTGTTGTTCAA	798		
Db	125	GCAAGCGGATGGCTCCCTCGTGACACATTTGCTCTCTATAAGCTCTTTACAAAAGCTTT	184		
QY	799	GTTTGTATTACAGTCTTGACATTTCTAGCAGCTATAGATGACGATTCAAGATAAAGTCGAT	858		

```

Query Match          1.9%; Score 44.8; DB 4; Length 1236;
Best Local Similarity 54.2%; Pred. NO. 0.00048;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY      1614 TGATTCAGAAGAGTAACTTCACTGTAATGTTCAGAACTTCAATGCTCTGTCCACATGT 1673
DB      696 TCAACCAATTAAATGACTATTACACAGCAGCTCTGGGACATCAATGGCAACTCCTCACGT 755

QY      1674 TAGCGGAATCACGTGCTTATTCGGTCTGCATACCCGAACTTGCTCTCCAGCTGCAATCAA 1733
DB      756 AGCTGTGATTGCGAGCCCTCTTTGTCCAAGCACACCCGAGCTGGACTCCAGACAAAGTAAA 815

QY      1734 ATCCGCAATTGATGACACACAGCGGATTGTACGATCGTCAAGGGAAGC 1781
DB      816 AACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAAATAGC 863

RESULT 4
US-08-894-818B-4
; Sequence 4, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/894,818B
  FILING DATE: 20-MAY-1998
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/JP96/03253
    FILING DATE: 07-NOV-1996
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: JP 323285/1995
    FILING DATE: 12-DEC-1995
  ATTORNEY/AGENT INFORMATION:
    NAME: Browdy, Roger L.
    REGISTRATION NUMBER: 25,618
    REFERENCE/DOCKET NUMBER: TAKAKURA=1
    TELEPHONE: (202) 628-5197
    TELEFAX: (202) 737-3528
  INFORMATION FOR SEQ ID NO: 4:
    LENGTH: 1566 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
    MOLECULE TYPE: genomic DNA
  FEATURE:
    OTHER INFORMATION: /note= N at position 1283 is G or T.
US-08-894-818B-4

Query Match      1.9%; Score 44.8; DB 3; Length 1566;
Best Local Similarity 54.2%; Pred. No. 0.00058;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1614 TGATTCAAGAAGAGTTAACTTCACTGTAATGTACAGAACTTCAATGTTGTGCCACATGT 1673
DB 696 TCACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCAACTCTTCACGT 755

QY 1674 TAGCGGAATCACTGCTCTTATCCGGTCTGCATACCCGAACTGGTCTCCAGCTGCATCAA 1733
DB 756 AGCTGGTATTGCAGCCCTCTTGTCTCCAAAGCACACCCGAGCTGGACTCCAGACAAAGTAAA 815

QY 1734 ATCCGATTCATGACACAGCGGATTTGTACGATCGTCAAGGGAAGC 1781
DB 816 AACAGCCCTCATAGAACTCTGATATCGTAAAGCCAGATGAATAGC 863

RESULT 5
US-08-894-818B-34
; Sequence 34, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/894,818B
  FILING DATE: 20-MAY-1998
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/JP96/03253
    FILING DATE: 07-NOV-1996
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: JP 323285/1995
    FILING DATE: 12-DEC-1995
  ATTORNEY/AGENT INFORMATION:
    NAME: Browdy, Roger L.
    REGISTRATION NUMBER: 25,618
    REFERENCE/DOCKET NUMBER: TAKAKURA=1
    TELEPHONE: (202) 628-5197
    TELEFAX: (202) 737-3528
  INFORMATION FOR SEQ ID NO: 34:
    LENGTH: 1962 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
    MOLECULE TYPE: genomic DNA
  FEATURE:
    OTHER INFORMATION:
US-08-894-818B-34

Query Match      1.9%; Score 44.8; DB 3; Length 1962;
Best Local Similarity 54.2%; Pred. No. 0.00068;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1614 TGATTCAAGAAGAGTTAACTTCACTGTAATGTACAGAACTTCAATGTTGTGCCACATGT 1673
DB 1092 TCACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCAACTCTTCACGT 1151

QY 1674 TAGCGGAATCACTGCTCTTATCCGGTCTGCATACCCGAACTGGTCTCCAGCTGCATCAA 1733
DB 1152 AGCTGGTATTGCAGCCCTCTTGTCTCCAAAGCACACCCGAGCTGGACTCCAGACAAAGTAAA 1211

QY 1734 ATCCGATTCATGACACAGCGGATTTGTACGATCGTCAAGGGAAGC 1781
DB 1212 AACAGCCCTCATAGAACTCTGATATCGTAAAGCCAGATGAATAGC 1259

RESULT 6
US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15

Query Match      1.9%; Score 44.8; DB 4; Length 1962;
Best Local Similarity 54.2%; Pred. No. 0.00068;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
```

QY 1614 TGATTCAGAGAGTAACTTCACTGTAATGTCAGGAACCTCAATGCTTCTCCACATGT 1673
Db 1092 TCAACCAATTAATGACTATTACACAGCAGCTCTGGACATCAATGGCACTCTCAGT 1151
QY 1674 TAGCGGAATCACTGCTCTTATCCGGTCTGCATACCCGAACCTGGTCTCCAGTGCATCAA 1733
Db 1152 ACTGGTATGAGCCCTCTTCTCCAGCACACCCGAGCTGGACTCCAGACAAGTAA 1211
QY 1734 ATCCGATTGATGACAGACGCGATTGTACGATCGTCAAGGGAAGC 1781
Db 1212 AACAGCCCTCATAGAACTGCTGATATCGTAAAGCCGATGAATAGC 1259

RESULT 7

US-09-313-294A-7407
; Sequence 7407, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1998-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 7407
; LENGTH: 295
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6476212 700381679H1
; NAME/KEY: unsure
; LOCATION: 4, 282
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-7407

Query Match 1.9%; Score 44.4; DB 4; Length 295;
Best Local Similarity 59.5%; Pred. No. 0.0023;
Matches 75; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 1633 TTCACGTAACTCAGGAACCTCAATGCTTGTCCACATGTTAGCGGAATCACTGCTTT 1692
Db 121 TTTCGAATGTTTCTGGAACCTAGCATGGCTGGCGCATATTGCTGGCAATTGCGCACTG 180
QY 1693 ATCCGCTCTGCATACCGCACTGCTCCAGCTGCAATCAATCCGATTCGATGACACAA 1752
Db 181 ATAAAGCAGAGAACCCAAAGTGGAGCCCTCAGCAATAAAATCTGCCCTGATGACACG 240
QY 1753 GCGGAT 1758
Db 241 GCGAAT 246

RESULT 8

US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491

; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc.feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 1.6%; Score 37.2; DB 3; Length 289;
Best Local Similarity 4.7%; Pred. No. 0.053;
Matches 11; Conservative 105; Mismatches 120; Indels 0; Gaps 0;
QY 1093 GGAGACTCATGTCATCCGGGAAAGGTATAAAGATCCGGGAGAGAGGTTGAGGTGATT 1152
Db 4 GGRACRARARURURARUR 63
QY 1153 TAGCTCACAGGAGGATAAAGAGTGTCTCTTTGAGAGGCTCACTTCCAAAGAGAA 1212
Db 64 NNRSNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNR 123
QY 1213 GAAATCCGAGCAAAATGTGATTTGTGATCGCGGAGTCAATGGAGATCGGAGAAAGGA 1272
Db 124 NNRSNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNR 183
QY 1273 GAAGCGGTTAAAGAGCTGGAGAGCTTGAATGATCTTAGCCCAATACAGAGATCAA 1328
Db 184 NNRSNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNRNNR 239

RESULT 9

US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc.feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 1.6%; Score 37.2; DB 3; Length 289;
Best Local Similarity 4.7%; Pred. No. 0.053;
Matches 11; Conservative 105; Mismatches 120; Indels 0; Gaps 0;

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,669A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

Query Match 1.6%; Score 37.2; DB 3; Length 32207;
Best Local Similarity 44.2%; Pred. No. 1.7;
Matches 153; Conservative 0; Mismatches 193; Indels 0; Gaps 0;
QY 961 GCAGCTGGTAAACAGCGTCCAAATCGAAAGCTCTGTTCGAAACACACAGCTCTCTGGTCTCA 1020
Db 19753 GCAGCAGGATGAACAGGAGCAGCAGGAGGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAG 19694
QY 1021 ACCATTGGCGAGGACACGCTTTGATCGAAGATTTCCCGCTGTGTCAGATTAGCCAAACGGA 1080
Db 19693 GTTAGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19634
QY 1081 AAGCTTCTCTATGGAGAGTCATTGTATCCGGGAAAAGGTATAAAGAAATGCGGAGGAGGAG 1140
Db 19633 GGAGTTAGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19574
QY 1141 GTTAGAGGATTTACGTACAGGAGGAGATAAAGAAATGCGGAGGAGGAGGAGGAGGAGGAGGAG 1200
Db 19573 GCAGGAGTTAGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19514
QY 1201 GTTAGAGGATTTACGTACAGGAGGAGATAAAGAAATGCGGAGGAGGAGGAGGAGGAGGAGGAG 1260
Db 19513 GCAGGAGTTAGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19454
QY 1261 TCGGAGAAAGGAGAGCGGTTAAAGAACTGAGAGGAGTTGCAATGA 1306
Db 19453 GCAGGAGTTGGAAGCAAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19408

RESULT 15

US-09-230-371A-20/c
Sequence 20, Application US/09230371A
Patent No. 6348586
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A
APPLICANT: Russo, James J
APPLICANT: Edelman, Isidore S
APPLICANT: Moore, Patrick S
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
FILE REFERENCE: 45185-G-PCT-US
CURRENT APPLICATION NUMBER: US/09/230,371A
PRIOR FILING DATE: 1999-11-17
PRIOR APPLICATION NUMBER: PCT/US97/13346
PRIOR FILING DATE: 1997-07-22
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
LENGTH: 32207

TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20
Query Match 1.6%; Score 37.2; DB 4; Length 32207;
Best Local Similarity 44.2%; Pred. No. 1.7;
Matches 153; Conservative 0; Mismatches 193; Indels 0; Gaps 0;
QY 961 GCAGCTGGTAAACAGCGTCCAAATCGAAAGCTCTGTTCGAAACACACAGCTCTCTGGTCTCA 1020
Db 19753 GCAGCAGGATGAACAGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19694
QY 1021 ACCATTGGCGAGGACACGCTTTGATCGAAGATTTCCCGCTGTGTCAGATTAGCCAAACGGA 1080
Db 19693 GTTAGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19634
QY 1081 AAGCTTCTCTATGGAGAGTCATTGTATCCGGGAAAAGGTATAAAGAAATGCGGAGGAGGAG 1140
Db 19633 GGAGTTAGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19574
QY 1141 GTTAGAGGATTTACGTACAGGAGGAGATAAAGAAATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
Db 19573 GCAGGAGTTAGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19514
QY 1201 GTTAGAGGATTTACGTACAGGAGGAGATAAAGAAATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
Db 19513 GCAGGAGTTAGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19454
QY 1261 TCGGAGAAAGGAGAGCGGTTAAAGAACTGAGAGGAGTTGCAATGA 1306
Db 19453 GCAGGAGTTGGAAGCAAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 19408

Search completed: February 3, 2004, 00:26:11
Job time : 122 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 23:29:01 ; Search time 623 Seconds
(without alignments)
13618.824 Million cell updates/sec

Title: US-09-806-767-1
Perfect score: 2328
Sequence: 1 atggaacccaacacctttt.....taacottgaagactaactga 2328

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	481.8	20.7	2265	10	US-09-938-842A-1065 Sequence 1065, Ap
2	481.8	20.7	2265	12	US-09-938-842A-1065 Sequence 1065, Ap
3	398.2	17.1	2295	10	US-09-938-842A-1186 Sequence 1186, Ap
4	398.2	17.1	2295	12	US-09-938-842A-1186 Sequence 1186, Ap
5	283.4	12.2	2043	12	US-10-260-238-1206 Sequence 1206, Ap
6	211.6	9.1	2073	12	US-10-260-238-485 Sequence 485, Ap
7	174.6	7.5	740	9	US-09-770-149-59 Sequence 59, Appl
8	153.2	6.6	606	12	US-10-260-238-5684 Sequence 5684, Ap
9	141.4	6.1	1467	12	US-10-260-238-703 Sequence 703, Ap
10	121.2	5.2	2334	12	US-10-260-238-1133 Sequence 1133, Ap
11	103.2	4.4	510	11	US-09-770-961-228 Sequence 228, Ap
12	100.2	4.3	2958	10	US-09-938-842A-338 Sequence 338, Ap
13	100.2	4.3	2958	12	US-09-938-842A-338 Sequence 338, Ap
14	98.4	4.2	849	12	US-10-260-238-3023 Sequence 3023, Ap
15	92.4	4.0	828	12	US-10-260-238-3022 Sequence 3022, Ap

C	16	90.8	3.9	513509	11	US-09-754-853A-4	Sequence 4, Appli
	17	90.2	3.9	447	12	US-10-260-238-5950	Sequence 5950, Ap
	18	88.8	3.8	2205	13	US-10-259-165-271	Sequence 271, App
	19	88	3.8	606	12	US-10-260-238-3101	Sequence 3101, App
	20	86.8	3.7	475	12	US-10-341-961A-346	Sequence 346, App
	21	83.4	3.6	363	10	US-09-878-574-2469	Sequence 2469, Ap
	22	79.6	3.4	1934	12	US-10-260-238-4334	Sequence 4334, Ap
	23	78.8	3.4	343	10	US-09-878-574-3844	Sequence 3844, Ap
	24	71.2	3.1	279	10	US-09-878-574-5412	Sequence 5412, Ap
	25	67.8	2.9	401	10	US-09-878-574-1414	Sequence 1414, Ap
	26	65	2.8	513509	11	US-09-754-853A-4	Sequence 4, Appli
	27	64.4	2.8	660	12	US-10-260-238-5484	Sequence 5484, Ap
	28	62.4	2.7	297	9	US-09-294-093B-6127	Sequence 6127, Ap
	29	62.4	2.7	379	10	US-09-878-574-2378	Sequence 2378, Ap
	30	59.4	2.6	1131	12	US-10-260-238-620	Sequence 620, App
	31	55.8	2.4	2451	10	US-09-938-842A-64	Sequence 64, Appl
	32	55.8	2.4	2451	12	US-09-938-842A-64	Sequence 64, Appl
	33	51	2.2	255	10	US-09-878-574-6446	Sequence 6446, Ap
	34	49	2.1	273	9	US-09-294-093B-2452	Sequence 2452, Ap
	35	48.4	2.1	2018	12	US-10-260-238-901	Sequence 901, App
	36	45.8	2.0	1049	13	US-10-140-472-358	Sequence 358, App
	37	45.8	2.0	1049	13	US-10-141-761-358	Sequence 358, App
	38	45.8	2.0	1049	13	US-10-142-885-358	Sequence 358, App
	39	45.8	2.0	1049	13	US-10-158-790-358	Sequence 358, App
	40	45.8	2.0	1049	13	US-10-137-871-358	Sequence 358, App
	41	45.8	2.0	1049	13	US-10-140-805-358	Sequence 358, App
	42	45.8	2.0	1049	13	US-10-140-864-358	Sequence 358, App
	43	45.8	2.0	1049	13	US-10-140-923-358	Sequence 358, App
	44	45.8	2.0	1049	13	US-10-141-756-358	Sequence 358, App
	45	45.8	2.0	1049	13	US-10-141-759-358	Sequence 358, App

ALIGNMENTS

RESULT 1

US-09-938-842A-1065
; Sequence 1065, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1065
; LENGTH: 2265
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1065

Query Match 20.7%; Score 481.8; DB 10; Length 2265;
Best Local Similarity 55.3%; Pred. No. 1e-141;
Matches 1033; Conservative 0; Mismatches 812; Indels 24; Gaps 4;

QY	380	ACTCGGCTATGGTCTAAATCTCGGTTGGTGGCAAGCACATTCGCGGCTTGATA	439
Db	344	AATCGGTTCCAGATCTCGGTTCTTCTTAACGGGCTTATCATCGGAGTTTATATA	403
QY	440	CTGGAGTTTGGCTGAAAGTCTTAGCTTTGACATACCGGAATCGCTTCGATCCACGA	499
Db	404	CTGGGTTTGGCTGAATCTAGAGTTCTGATGATGATGCTGAGATTCCTCTTA	463

ORGANISM: Arabidopsis thaliana
US-09-838-842A-1065

Query Match 20.7%; Score 481.8; DB 12; Length 2265;

Best Local Similarity 55.3%; Pred. No. 1e-141;

Matches 1033; Conservative 0; Mismatches 812; Indels 24; Gaps 4;

```
QY 380 ACTCCGCTGATGGTCTAAATCTCGGTTTGGTCAAGGCACAAATATCGGGTGCCTTGATA 439
DB 344 AATTCGGTGTACGATCTCGGTTCTTCTTAACGGCGTTATCATCGGAGTTTATAGATA 403
QY 440 CTGGAGTTTGGCCTGAAGTCTTAGCTTTGACGATACCGAATGCCCTTGATTCACCGGA 499
DB 404 CTGGCGTTTGGCCTGAATCTAGAGCTTCGATGATCTGATATGCTGATGATTCCTCTCA 463
QY 500 AATCGAAGGGAATTTGCCAAGAGGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAAGC 559
DB 464 AATGGAAGGAGATGTGAATCTGTTCCGATTTCCGATTCGATTTCCGATTTCCCTCA 523
QY 560 TAATCGGTGTAGATTTCTCATCAGAGGACACCGTGTGCTTAATTCACGAGGAATCAC 619
DB 524 TTATCGGAGCTAGAAGCTTCTCAAAGGATTTCAAATGGCTTCTGGTGGTGGTTTTCGA 583
QY 620 CAACATGCTCGTGAATATATTCGCAAGAGATTCACGGGACACGGGACTCACACCG 679
DB 584 GTAA-----GCGTGAATCTGTTCTCTCGTGTATGACGACATGGAACACATCTT 637
QY 680 CCTCAACAGTTGGTGGATCTCTGTTTCGATGGCAATGTTCTTGGCAATGGAGCTGGTG 739
DB 638 CAACTACCGCGGGGATCCGCGGTAGAAACGCTAGCTTCTTCGGTTACGCCCGCGGTA 697
QY 740 TGGCTCGTGGATGGCTCTCGGAGCTCACATTCAGTCTATAAAGTCTGTGGTTCAATG 799
DB 698 CGGCCAGAGTATGGCCACTCGTCTCGTGTGCTACTTATAAAGTTTGTGGAGTACTG 757
QY 800 GTTGTTACAGCTCTGACATCTAGCAGCTATAGATGAGCATCAAGATAAAGTGCATG 859
DB 758 GTTCTTTTGGATCTGATATAGTCTGCTATGATCGAGTATCTTGTGGTGGTGGTGGT 817
QY 860 TTCTTTCCGCTTCCCTGGCGGTTCCCTATTCTTTGATGATGACAAATCGCCATTG 919
DB 818 TGCTTTCTGTTATCTCTTGGTGGTGGTCTCTCGTATTAATCGGATAGCATTTGCCATTG 877
QY 920 GAACATTCGAGCCATGAAACGGGTATATCTGTAACTCTGAGCTGTGTAACAAACGGTTC 979
DB 878 GAGGTTTTCGGCTATCGAGAGAGGTTGTTTGTGCTGTTGTTCTGCTGTTGATAGTGTC 937
QY 980 CAATCGAAAGCTCTGTTGCAACACAGCTCTCTGGGTCTCAACCATTTGGCGCAGGCAAGC 1039
DB 938 CTACTAGAGCTTCTGTTGCCAATGTTGCTCTCTGGGTATGACTGTTGGTGTGTTACTT 997
QY 1040 TTGATCGAAGATTTCCGCTGTGGTCAATAGCAGGAAAGCTTCTATGGAGAGT 1099
DB 998 TAGATAGAGATTTTCCGGCTTTTCGAACTCTCGTAAACGGGAAACGACTTACCGGTGTTT 1057
QY 1100 CATTTGATCCGGGAAAGGTATAAGATCCCGGAGAGAGTTGAGGTGATTTACGTCA 1159
DB 1058 CGCTGTATAGCGGTAGGAATGGGA-----CGAAGCCGCTTGAATTTGTTTATAATA 1111
QY 1160 CAGGAGGAGATTAAGGAAGTCAAGTTCTGTTTGAAGAGGTCACATTCGAAGAGAAATCC 1219
DB 1112 AAGGGAATAGTATCGAGTAATCTTGTTTACTGTTCTGTTGATTCAGTATTGTTTC 1171
QY 1220 GAGCAAAATGGTATTTGATCGCGAGTCAATGAAGATCGGAAGAGGAGGAGCGG 1279
DB 1172 GTGGGAAGATGTTGTTTGTGATAGAGGTGTTAATGCTAGAGTTGGAAGAGGAGCTGTGG 1231
QY 1280 TTAAGAGAGCTGGAGGAGTTCATGATCTTAGCCATACAGAGATCAACCAAGAGAG 1339
DB 1232 TTAGAGATGCTGGTGGTTTGGGATGATAATGCGAATCTGCTGCGAGTGGAGAGAGC 1291
QY 1340 ATTCTATTGAGTTCTCTTACCAGCTACATTTGTTGTTTACACTGATGAGTCTCTTC 1399
```

```
DB 1292 TTGTGGCGATAGTCAATTTGCTTCCCGGATCGCTGTAGGAAGAGACTGGTGAATTAC 1351
QY 1400 TGAAGGCTTATGTTAATGCGACCGGTGAAA CAAAGGCGCGGATAAATTTTGGTGGTAGCG 1459
DB 1352 TTAGGAGTATGTTAAGTCAATTTAAACCAACCGCTCTTCTTGTGTTTAAAGAAACGG 1411
QY 1460 TGATTTGGAGTCAAGGACCGGAGGTGGCTAGTTTTCAGCTCGAGGACCGAGTTTAG 1519
DB 1412 TTCTTGAGGTTAAGCGGTTCTCTGTTGTTGCTTTTAGCTCGAGAGGTCCTAATACTG 1471
QY 1520 CCAATCCCTTCGATATAAACCAGGATATGTTGCTCCGGGAGTCAATATCATTCGCGGTT 1579
DB 1472 TTACTCTGAATCTTGAGGCTGATGTTATTTGCTCTGGAGTTAATATTTTGGCTGGTT 1531
QY 1580 GGCCTCAAAATCTAGGACCAACCGGACTTCTTTATGATTCAGAGAGTAACTTCACTG 1639
DB 1532 GGTCTGACGCTATTGGTCTTACTGCTTGAAGAGTCTTAGGAGGACTCAGTTCAACA 1591
QY 1640 TAAATGTCAGGAACCTCAATGCTTGTCCAGATGTTAGGGGAATCAGTCTCTTATCCGGT 1699
DB 1592 TCATGTCAGGTACGTCATATGTCATGCCACACATCAGTGGTTTAGCGGGTCTTTTGAAG 1651
QY 1700 CTGCATACCCGAATCTGGTCTCCAGCTGCAATCAAAATCGCATTCGATGACACACGCGATT 1759
DB 1652 CAGCTCACCTGAGTGGAGTCCGAGTGTATCAAAATCAGCTCTCATGACTACAGCTTACG 1711
QY 1760 TGTACGATCGTCAAGGAAACGATAAAGGATGGT-----AACAAACCCAGCGGTGTG 1813
DB 1712 TTCCTGACAAACCAACGCTCTCTCCATGATGCTGCAGACACACAGCTATCTAACCCAT 1771
QY 1814 TTGGATTTGGAGCAGGCGATGTGAATTCGGAAGGAGGATAAACCCGGGATGTTGTTACA 1873
DB 1772 ATGCTCAGGCTCGGGCCATGTAGATCCCAAAAGGCTCTCTCACCAGGCTTGTGTACG 1831
QY 1874 ACATTCACCACTGGATGATTAATACTTACCTCTGCACTCTTGGATTACAAAGATCAGATA 1933
DB 1832 ACATCTCAACCGAGGATACATCAGGTTTTTGTCTCTCTAGACTACACAGTCGATCACA 1891
QY 1934 TTTTAGCAATCACTATAAGAACGTCGAGCTGCAATGGAATATTTCGGGAAAAACCCGGGTT 1993
DB 1892 TTGTTGCAAT-----TGTGAAGCGACCTAGCTTAAGTCTCGAAGAGTCTTCAGATC 1945
QY 1994 TTAGTCTCAATTAACCGTTCGATAGCCGTGATTTTCAAACGTGGCAAGACTACGAGATGA 2053
DB 1946 CTGGTCACTCAACTACCCAAAGTTTCTCGGTTTTTGTGGGGGTAAGAGTTGTGCGGT 2005
QY 2054 TCACAGCGGTGTCATTAACGTTGGAGTCTTAATCTGATATCTCATGTAATGTCAAGG 2113
DB 2006 ACATTCGGGAAGTAAACAAATGTTGGTGCAGCAAGCTCGGTTTACAAAGTCAAGGTTAATG 2065
QY 2114 CTCAGAGGGGATCAAGATTAATGCTCAATCTTAAGAGACTTGTCTTCAAAACGTTGGATC 2173
DB 2056 GAGCTCTCTAGTGTGGGATCTCTGTTAAACCATCGAACTTTTCGTTTAAAGCGTGGAG 2125
QY 2174 AGACCTCAGCTATAGAGTATGTTGTTGTTTGAAGAGAGAAAAACAGAGGAGGAGGTGG 2233
DB 2126 AGAAGAGAGGTACACAGTCAAGTTTGTACGAGAGAGGAGTATGATGACGACAGG 2185
QY 2234 CTAGCTTTG 2242
DB 2186 CTGAGTTG 2194
```

RESULT 3

US-09-938-842A-1186
; Sequence 1186, Application US/09938842A
; Patent No. US20020160378A1

GENERAL INFORMATION:
; APPLICANT: Harper, Jeff

; APPLICANT: Kreps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1186
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-1186

Query Match 17.1%; Score 398.2; DB 10; Length 2295;
Best Local Similarity 53.2%; Pred. No. 4.5e-115;
Matches 1072; Conservative 0; Mismatches 898; Indels 45; Gaps 9;
QY 205 TCTCGACTTCTTACTCCTATGCTCTGGATTGAAGGATTGCTGCTCAGTTGACTGAA 264
DB 166 TCTCGAATCTCCATGTTTACCACACAGTCTTCCAGTTTCTCCGCCGTTGTACTCCA 225
QY 265 TCAGAACCCGAGATACAGATATTCACCTGAAGTTGTTGCAGTGAGACCTGACCATGTT 324
DB 226 GATGAACAGATATCTCCGTACCCAGCCAGAGTCTTGTCTTTTCGAGACCGACGT 285
QY 325 CTTTCAGGTTCAAAACCACTTACTCTTACAAAGTTCTTGGAGTCCGAGTTTGGAAACTCC 384
DB 286 CGAGAGCTTCACACCACACGTTCTCTCAATTTCTTGGTTTACAA-----AACCAAAA 339
QY 385 GGTGTATGTCTAAATCTCGTTTGGTCAAGGACAAATATTCGGCTGCTTGNATCTGA 444
DB 340 GGAATATGTGTGAGATCTGATTAAGGATCAGAGCTAATCATTTGGCGTTTTCGACCGGA 399
QY 445 GTTTGGCTGAAAGTCTAGCTTTGAGTACCGGAATGCTTTCGATTCCACGGAATGG 504
DB 400 ATTTCGGCGGAGCGAGGATTTCTCAGATCTTAACTCGTCCCAATTCGAAAGGTGG 459
QY 505 AAAGGATTTCCGAAGAGAGAAAGTTTCAGTTCTTCGAGCTGTAAACCGGAAGTAAATC 564
DB 460 AGAGCGTTTCGGAATCCGGAGCGAGATTTCAGTCTCGGAATGTAAACCGTAAATATC 519
QY 565 GGTGCTAGATTTTCATCAGAGACACCGTGTGCTTAATTCACAGAGGAATCACCAAC 624
DB 520 GGAGCAAGATTTCTCGCTAAGGGACAAACAGCGCT--GTAATCGGAGGAATCAACAAA- 576
QY 625 ATGCTCTGTAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACCGGCTCA 684
DB 577 ---ACCGTTGAGTTTCTATCTCTCTGAGCGCGATGACACGGTACTCACATTTCTCTCA 633
QY 685 ACAGTTGGTGGATCCTCTGTTTCATGCGGAATGTTCTTCGCAATGAGCTGTGTGGCT 744
DB 634 ACCGCCCTGGCGGTACGCTTTTAAAGCAGATATGTCGGTTACGCTTCGCGTGTAGCC 693
QY 745 CGTGGATGCTCTCGAGCTCAATTCGAGTCTATAAAGTCTGTGGTTCAAT---GGT 801
DB 694 AAAGTGTGTCTCCAAAAGCTCGTATCGCGCCTACAAAGTCTGTGGAAAGATTCCGGT 753
QY 802 TGTACAGCTCTGACATTTAGCAGCTATAGATGATAGATTCAGATAAAGTCGATGTT 861
DB 754 TGTCTGATTCGATATTTCTGCGCGCTTTGATGCGCGCTTTGATGCGCGCTTTAGAGACGGTGTGCGACGT 813
QY 862 CTTTCGGTTTCCCTTGGCGTT-----TCCTATTCTTTGTTATGATGACAAATC 912
DB 814 ATATCGATCTCAATCGTGTGGAGCGGATTAATTCGCGGTATTAACCTCGATCCAATC 873
QY 913 GCCATTGGAAATTCGAGCCATGGAACCGCGGTATATCTGTAATCTGTGAGCTGTGAAC 972
DB 874 GCTATAGGCTGTACGGCGCGCGTTCGAAAGGAATCTTCGCTCTTCTCTCTGCGGGAAC 933

QY 973 AACGGTCCAATCGAAGCTCTGTTGCACACACAGCTCTTGGGTCTCAACCATTTGGCGCA 1032
DB 934 GAAGGACCTAACCGTATGTTCAGTATCAGAACCTCCGCGCGTGGTGAACACCGTTGGTGT 993
QY 1033 GGCACGCTTGATCGAAGATTTCCCGCTGTGGTCAGATTAGCCAAACGGAAGGCTTCTCTAT 1092
DB 994 AGTACAAATCGATCGGAATTTCCAGCGGATGCTATTCTCGCGACGACATCTCTCAGA 1053
QY 1093 GGAGAGTCATTGTATCCGGGAAAAGGTATAAAGAAATCCCGGAGAGAGGTTGAGGTGATT 1152
DB 1054 GGAGTGTCTCTTTACGCTGGAGTACCTTTAAACGGTCGTATGTTTCCGGT---GGTTTAT 1110
QY 1153 TACGTACAGAGGAGATATAAGAAAGTGAATTCGTTTGAGAGGGTCACTTCCAAGAGAA 1212
DB 1111 CCCGGTAAATCGGGAATGTCTATCAGCGCTCTCTATGATGAGAAACACGCTTGATCCGAAG 1170
QY 1213 GAAATCCGAGCAAAATGGTGATTTGTATCGGGAGTCAATGGAAGATCGGAAAGGA 1272
DB 1171 CAAGTAGGGGTAAATAGTAATCTGCGATAGAGGAAGCAGTCCACGCGTAGCCAAAGGA 1230
QY 1273 GAAGCGGTTAAAGAAAGCTGGAGGAGTTGCAATGATCTTAGCCCAATACAGAGATCAACCAA 1332
DB 1231 TTGGTTGTGAAGAAAGCAGGTGGTGTGCGAATGATTTCTCGCTAATGGAGCATCTAACGGT 1290
QY 1333 GAAGAAGATCTATTGACGTTTATCTTACAGCTACATTTGATTTGTTTACATGATGATCA 1392
DB 1291 GAAGGATTTAGTCGAGATGCTCATCTTATTCAGGCTGTGCGGTTGGATCAACAGGAAGA 1350
QY 1393 GTCTCTTGAAGGCTTATGTTAATGCCAGGTGAAACCAAGCGCGGATAATTTTGGT 1452
DB 1351 GATAGAAATCAAGCATATGCTTCTTACATCCGAATCCAATTTGCTTCAATTTGATTTTCTCAGA 1410
QY 1453 GGTACGCTGATTGGAGGTCACAGAGCACCGAGGTGCTCAGTTTTCAGCTTCGAGGACCG 1512
DB 1411 GGAATCTAGTTGGATTAACCGGCTCCGTTATGCTTCTTCTTCTCCGCTAGAGGACCA 1470
QY 1513 AGTTTACCAATCTCTCGATACATAAACCAGGATATGATTCCTCGGAGTCAATATCATTT 1572
DB 1471 AACGTTTAAAGCCCGAGATTTCTTAAACCGGATTTGATTTGCTCCCGAGTTAATCCTC 1530
QY 1573 CGGGTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTTCAAGAGAGTTAATC 1632
DB 1531 CCCGATGGACAGACGCTGTTGGACCTTACAGGTTTGGCGTCAGATCCAGGAAACCGAA 1590
QY 1633 TTCACTGTAATGTCAGGAATCTCAATGTTTGTCCACATGTTAGCGGAATCCTCTGCTCTT 1692
DB 1591 TTCAACTTCTCTCCGTAATCTTCAATGGCATGCTCTCAGTTAGTGTGCGGCGCGCTT 1650
QY 1693 ATCCGCTCTGCATACCCGAACTGCTCTCCAGCTGCAATCAAAATCCGCAATGATGACAAAC 1752
DB 1651 CTCAATCCGCTCATCCAGATTGGAGCCCTCCGCTGATACGATCGGCAATGATGACAAACG 1710
QY 1753 GCGGATTTGTACGATTCGTCAGGGAAGCGGATAAAGGATGTTAATC-----AAACGAGCC 1806
DB 1711 ACTAACCTCGTCGATAACTCTAACCGCTCGTTGATCGATGAATCCACCGGGAATCCGCT 1770
QY 1807 GGTGTGTTTCGATTGGAGCAGGCGATGTAATCCGCAAAAGGCGGATAAACCCEGATG 1866
DB 1771 ACGCCCTTATGATTACGGGTCCGGTCATTTAAATTTGGCGCGGCTATGAATCCCGGCTT 1830
QY 1867 GTTTACAACTTAAACAGTGGATTATCAATCTTACCTCTGCACTCTTGGATTTCACAGA 1926
DB 1831 GTCTACGATATACTAATGATGATTTACATTCAGTTTCTTGTCTCCATCGGTACGGACCA 1890
QY 1927 TCAGATATTTTAGCAATCCTCATTAAGAACTGAGCTGCAATGGAATTTTCGGGAAAC 1986
DB 1891 AAGACGATCCAAAGTGATAACAAAGAACCCGCTGAGATG---TCGAGCGACGAGGAAACCG 1947
QY 1987 CGGGTTTTAGTCTCAATTAACCGTCGATACCGGTGATTTC-----AAACGTTGGC 2037
DB 1948 TCTCCCGGGAATTTGAATTTATCTTCGATACCGCGGTGTTTCTTACTTAATAGAGAGGA 2007
QY 2038 AAGACTACGGAGATGATCACAAGGCGTGTCTACTAACGTTGGGAGTCTCTTAACCTCGATATAC 2097

Db 2008 TTGGTGAATAAACTGTTATAAGACCGCGACGATGTCGGCAGGCTGAGCGGTTAT 2067
Qy 2098 TCAGTGAATGCAAGCTCCAGAGGGATCAAAATGTTCAATCTAAGAGACTGTG 2157
Db 2068 CGGCGAGGATAGAGTCCGCGAGAGGAGTACCGGTGACAGTGAACACCACTAGCGTTG 2127
Qy 2158 TTCAACACGTTGGATCAGACCGCTGAGCTATAGAT 2192
Db 2128 TTACTTCGGCGTTAAGAGACGGAGCTATCGGT 2162

RESULT 4

US-09-938-842A-1186
; Sequence 1186, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1186
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1186

Query Match 17.1%; Score 398.2; DB 12; Length 2295;
Best Local Similarity 53.2%; Pred. No. 4.5e-115;
Matches 1072; Conservative 0; Mismatches 898; Indels 45; Gaps 9;
Qy 205 TCTCGACTTCTACTCTCTATGGCTCTGCGATTGAAGGATTGCTGCTCAGTTGACTGAA 264
Db 166 TCTCGAATCGTCCATGTTTACCACACAGTCTTCCATGTTTCTCGCGGTTGTTACTCCA 225
Qy 265 TCAGAACCGGAGATCTAGATATTCACCTGAAGTTGTCAGTGAGACCTGACCATGTT 324
Db 226 GATGAAGCAGATAATCTCCGTAAACCCACGACGTTCTGCTGTTTTCGAAGACCGAGT 285
Qy 325 CTTGAGGTTCAACACATTAATCTTACAGTTCTTGGACTCGACGGTTTGGAACTCC 384
Db 286 CGAGAGCTTCAACACACAGTTCTCTCTCAATTTCTTGTTTACAA-----AACCAAAA 339
Qy 385 GGTGTATGCTCTAAATCTCGGTTTGGTCAAGGCAATATTCGCGGTGCTGTGACTGGA 444
Db 340 GGACTATGCTCAGATCTGATACGGATCAGACGTAATCATGTCGCTTTTGACACCGGA 399
Qy 445 GTTTGGCTGAAAGTCTTAGCTTTTACGATACCGGAATGCTTCGATTCACCGGAATGG 504
Db 400 ATTTGGCCGAGCGGAGGAGTTTCTCAGATCTTAACCTCGGTCCCAATTCCAAAAAGGTG 459
Qy 505 AAAGGATTTGCGAAGAGGAGAAAGTTTCAAGTTCTTCGAGCTGTAAACCGGAGCTAATC 564
Db 460 AGAGCGTTTTCGATTCGAGCGGACGAGATTTCAGTCTCGGAATGTAACCGGTAAATATC 519
Qy 565 GGTGTAGATCTTTCATCAGAGGACACCGTGTCTGCTAATTCACAGAGGAAATCACCAAC 624
Db 520 GGAGCAAGATTTCTCGTTAAGGACCAACACCGCT--GTAAATCGGAGGAATCAACAA- 576
Qy 625 ATGCTCTGTAATCATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCGCTCA 684

Db 577 ---ACGGTTGAGTTTCTATCTCTCGTGACGCCGATGCACACGGTACTCACACTTCTCTCA 633
Qy 685 ACAGTTGGTGATCTCTGTTTTCGATGCGGAATGTTCTTGGCAATGGAGCTGGTGTGCT 744
Db 634 ACCGCCGTGGCGTCAAGCTTTTAAAGCGAGTATGTCGGTTACGCCCTCGGTTAGCC 693
Qy 745 CGTGGATGGCTCCCTGGAGCTCACATTGCACTATATAAGTCTGTTGTTGTTCAAT---GGT 801
Db 694 AAAGGTGTTGCTCCAAAGCTGATCGCGCTACAAAGTCTGTTTGAAGATTCGGGT 753
Qy 802 TGTTCAGCTCTGACATTTAGCAGCTATAGATGATGAGATTCAGATTAAGATTCAGATTT 861
Db 754 TGTCTCGATTCCGATATTTCTCGCCGCTTGTATGCGCTGTTAGAGACGGTGTGACGTT 813
Qy 862 CTTTCGCTTTCCCTTTGGCGGTT-----TCCCTATTCCTTTGATGACACAATC 912
Db 814 ATATCGATCTCAATCGGTGGTGGAGACGGGATTTACTTCGCGGTATTAATCTCGATCAAATC 873
Qy 913 GCCATTGGAACATTCGAGACCATGGAACGCGGTATATCTGTAATCTGTCAGCTGCTAAC 972
Db 874 GCTATAGGCTCTGACGCGCGCGTGAAGGAATCTTCGCTCTCTCTCTGCGCGGAAC 933
Qy 973 AACGGTCCAAATCGAAAGCTCTGTTGCAACACAGCTCCTTGGGTCTCAACATTCGGCA 1032
Db 934 GAAGACCTAACGGTATGTGATGACGAACTCGCGCGTGGTAAACCAACCGTTGGTGT 993
Qy 1033 GGCACGCTTGATCGAAGATTTCCCGCTGTGTCAGATTAGCCAAACGGAAGCTTCTCTAT 1092
Db 994 AGTACAATCGATCGGAATTTCCACGCGATGCTATTTCTCGCGACGACATCGTCTCAGA 1053
Qy 1093 GGAGAGTCAATGTTATCCGGGAAAGGTATAAGAAATGCGGGAGAGAGGTTGAGGTGATT 1152
Db 1054 GGAGTGTCTCTTTACGCTGGAGTACCTTTAAACCGGTCTGATGTTTCCCGT---GGTTTAT 1110
Qy 1153 TACGTCACAGGAGGAGATAAAGGAAGTGAAGTCTCTGTTTGGAGGGTCACTTCCAAAGAA 1212
Db 1111 CCGGTAATTCGGGAATGTCATCAGCTCTCTATGATGGAGAACACGCTTATCGGAG 1170
Qy 1213 GAAATCCGAGGCAAAATGTTGATTCGCGAGTCAATGGAAGATCGGAGAAAGA 1272
Db 1171 CAAGTGAGGGTAAATAGTAACTCTCGATAGAGGAAGCAGTCCACGCTAGCCAAAGA 1230
Qy 1273 GAAGCGGTTAAGAGAGCTGGAGGATGCAATGATCTTAGCCAAATACAGAGATCAACCA 1332
Db 1231 TTGGTTGTGAAGAAAGCAGGTGGTGTGGAATGATTTCTCGCTAATGGAAGCATCTAACGT 1290
Qy 1333 GAAGAGATCTTATGACGTTTCACTCTTACAGCTACATGATTGTTGATCACTAGTCA 1392
Db 1291 GAAGGATTAGTCGAGATGCTCATCTTATTCAGGCTGTGCGGTTGGATCAACAGAGA 1350
Qy 1393 GTCTTCTGAAGGCTTATGTTAATGCCAGGTTGAACCAAGCGCGGATATTTTGGT 1452
Db 1351 GATAGATCAAGCATATGCTTCTTCAATCCGATCCAAATGCTTCAATGATTTCAGA 1410
Qy 1453 GGTACGCTGATTGGAGGTCACAGCAGCGGAGTGGCTCAGTTTTCAGCTCGAGGACCG 1512
Db 1411 GGAACATATAGTTGGATTAAACCGCTCGGTTATGCTTCTTCTCCGTTAGAGACCA 1470
Qy 1513 AGTTAGCCAAATCTTCGATATAAAGCGGATGATGTTGCTCCGGAGTCAATATCAT 1572
Db 1471 AACGGTTTAAAGCCCGGAGATTTTAAACCGGATTTGATTGCTCCCGGAGTTAACTCCTC 1530
Qy 1573 GCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTATGATTCAAGAGAGTTAAC 1632
Db 1531 GCGCATGACAGACGCTGTTGGACCTACAGGTTTGGCGTCAATCCAGGAAACCGAA 1590
Qy 1633 TTTCACTGTAATGTCAGGAATTTCAATGCTTCTCCACATGTTTGGCGGATCACTGCTCT 1692
Db 1591 TTCAACATTTCTCTCGGTACTTCAATGGCATGCTCTCAGTTAGTGTGTCGCGCGGCT 1650
Qy 1693 ATCCGCTGTGCATACCGGAATGCTTCCAGTGTCAATCAAAATCCGCAATGTATGACA 1752
Db 1651 CTCMAATCCGCTCATCCAGATTGGAGCCCTGCGGTGATACGATCGCAATGATGACAACG 1710

1727	QY	CAATCAAAATCCGCAATTGATGACAAACAGCGGATTTGTACGATCGTCAAGGGAAAGCGATAA	1786
1487	Db	CGATCAAGTCGCGCAATCATGACAAACAGCTTATGTGTTGATACCGAAGAACGCAATCC	1546
1787	QY	AGGATG---GTAACAAACAGCGGCTGTGTTGGGATTCGAGCAGGCGTGTGAATCCGC	1843
1547	Db	TGGATGAAGATACAATATTGCAAGGCGAATTCGCGAGTTGTGTGGACATGTTAGTCCTT	1606
1844	QY	AAAAGGCGATAAACCCCGGATTTGGTTTACAACATTCAACAGTGGATTACATAAATTTACC	1903
1607	Db	CTGAGCCATTGACCCCGCTGATCTATGACATTGATGATGCACAATACATTTCATATC	1666
1904	QY	CTGTCACTCTTTGGAATTCACAAGATTCAGATATTTTACAAATCACTCATAPAGACGTGAGCT	1963
1667	Db	TCCTGTGSCCTGGGTTACACAGATGTTTCAAGTTTCAGATAAATGGCAAAATCGAAAGATGCGCT	1726
1964	QY	GCAATGGAATATTGCGGAGAAAACCCGGTTTTAGTCTCAANTACCGCTCGATAGCCGTGA	2023
1727	Db	GCAAGGAA-----GCAAAATTAATCGAAGCTGAACCTTAACCTTCCATCAGTAGCAGTAA	1780
2024	QY	TTTTTCAAACGTGGCAAGACTACGAGAGATGATCAAGCGCTGTCACTAAACGTTGGGAGTC	2083
1781	Db	GGGCAAGTCTGGCAAGCTTTGTGTGAACCGGAC-----TGTACTAATGTGGGAGAAG	1834
2084	QY	CTAACTCGATATACTCAGTGAATGTCAAGGCTCCAGAGGGGATCAAAGTTATTGTCAATC	2143
1835	Db	CAAAATTCAAAGCTACACCGTGGAAATTGATATGCCTAGGGAGGTGATGACAAGTGTGTCCAC	1894
2144	QY	CTAAGACATTTGTTTCAAACAGCTGGATCAGACGCTGAGCTATAGAGTATGGTTTGTAT	2203
1895	Db	CAACAAAGCTAGAAATTCACCAAGATGAAGAGAGAGAAAACATTTTCATTAAGCTTAAGTT	1954
2204	QY	TGAAGAGAGAAAAACA	2218
1955	Db	GGGATATTAGCAAAA	1969

RESULT. T 6

US-10-260-238-485

03-10-200-230-403
: Sequence 485, Application US/10260238

Sequence 483, Application US/10
: Publication No. US20040016025A1

: GENERAL INFORMATION:

APPLICANT: Budworth. paul R

APPLICANT: Moughamer, Todd G

APPLICANT: Moughamer, Todd G

APPLICANT: BRIGGS, Steven P.
APPLICANT: CROOK, Brock
APPLICANT: CROOK, Brock

APPLICANT: Cooper, Bret
APPLICANT: Classbrook, Tino

APPLICANT: Glazebrook, Jane

APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyuki

APPLICANT: Kreps, Joel

APPLICANT: Provart, Nicholas

; APPLICANT: Ricke, Darrell

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: PROMOTERS FOR REGULAT

FILE REFERENCE: 60111-NP

; CURRENT APPLICATION NUMBER: US/10

CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/325,448

PRIOR FILING DATE: 2001-09-26

: PRIORITY APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/323,211

; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: IIS 60/370-620

; PRIOR APPLICATION NUMBER: US 60/370,620
 : PRIOR FILING DATE: 2002-04-04

; PRIOR FILING DATE: 2002-04-04
 ; NUMBER OF SEQ ID NOS: 6077

```

; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 195

```

; SEQ ID NO 485
; LENGTH: 2073

LENGTH: 2073

TYPE: DNA

Query Match 9.1%: score 211.6: DB 12: Length 2073:

Query Match 5.1%; SCORE 211.6; DB 1
Regt [local] Similarity 51.4%; Pred No 1 2e-55.

Matches	686;	Conservative	0;	Mismatches	609;	Indels	39;	Gaps	7;
622	AA	CATGCTCTGTCGAATACATATTTCCGCAAGAGATTCAACGGGACACGGGAGTCTCACACGCC		681					
367	AATG	GACACACAGAGTCCAAAGTCAACCTTCGACACCGAAAGCCACGCGCACACTGCA		426					
682	TCAA	GAGTTGGTGGATCCTCTGTTTCGATGGCGGAATGTTCTTGCAATGAGCTGGTG		741					
427	TCTAC	CGCGCTGGTTTCAGCGCTCCCTGGCGTAACTTTGTTGGCTATGCCAATGGCAACC		486					
742	GCTCGT	GGGATCGCTCTCGGAGCTCACATTCAGTCTATAAAGTCTCTGTTGTTCAATGGT		801					
487	GCCCA	AGGCATGGCTATAGTGCACACATTCGATCCCTACAAAGTATCTCGAGATGAT		546					
802	-----	TGTTACAGCTCTGACATCTTAGCAGCTATAGATGTAGCGATTCAAGAT		849					
547	GGAA	ATGCATCGTGGCTACCTTGACATCTCGCGGAATGAATGAGGCAATGCAGAT		606					
850	AAAGT	CGATGTTCTTTTCGTTTCCCTTGCGCGTTTCCCTATCTCTTTGTTATGATGACACA		909					
607	GGAGT	AGATGTATATCCCTTTCCCTTGGTGGTCTAAAGGCCACAACCTCTACAAGAACCC		666					
910	ATCGC	CAATGGAAATTCGAGCCATGGAACCGGTATATCTGTAATCTCTGCAGCTGGT		969					
667	ACCTC	CTTTGGCGCAATCAATGCCATCGTAGGCAATGTCGTGTTCCACTCCGCTGGA		726					
970	AACA	CGGTCCAAATCGAAGCTCTGTTGCAACACAGCTCTTGGGTCTCAACCAATGGC		1029					
727	AATG	ATGGCCCTGGCAGTATCTGCTAATTAACCTCGACCAATGGTGTATACGGTTGGC		786					
1030	GCAGG	CACGCTTGATCGAAGATTTCCCGCTGTGTCAGATTAGCCAACGGAAAGCTTCTC		1089					
787	GCAT	CAAGCATTTGACCGTCGGTTCCAGCTCAACGTTGCTTGGACATAATCGCGGAACC		846					
1090	TATGG	AGAGTCAATTGATTCGCGGAAAGGTATAAGAAATCCCGGAGAGAGGTAGGTTG		1149					
847	TACAT	CGGCACCTCG---CTGTATTTCGGGCCAAACACAGCCGGTTCATTTCACCTCTA		903					
1150	ATTTC	AGTCAAGGAGGAGATAAAGGAAGTCAAGTTCTGTTTGGAGAGGTCACTTCCAAGA		1209					
904	GTGTAC	-----GGTGGTATGCTGGTTCCTCTGTGTGAATACGGGATGCTTTCACGC		957					
1210	GAAGA	MAATCCGAGGCAAAATGGTGATTTGTATCGGAGTCAATGGAA-----GATCG		1263					
958	AACAT	GTCTACCGGAAGATGTTCTCTGCTATGCGCAAAAAATACTCAACCCCAATA		1017					
1264	GAGAA	AGGAGAGCGGTTAAAGAGCTGGAGAGTTGCAATGATCTTTAGCCATACAGAG		1323					
1018	GTCTA	AGAAGCTGCTGTCTCAACAAGCCGGTGGTGTGGTCCATCATTTCAATTGTTCCA		1077					
1324	ATCA	ACCAGAGAGATTCATTGACGTTTCATCTTACAGTCAATTGATTGGTTAC		1383					
1078	GAATA	TGGTATTTCTTAAGAGCTTCGCTGATATCTTGGCAACGTCGACGATTAACCTTC		1137					
1384	ACTG	AGTCACTCTCTGAAGCTTATGTTAATGCCACGGTGAACCAAGGCGCGGATA		1443					
1138	AAGCA	CACCGAAACGATCCATTCGTACACACAGTCGGTGGCCGACCGGTTGCAAGATC		1197					
1444	ATTTT	TGGTGGTACCGGTGATTGGGAGTGCAC---GAGCACCGAGGTGGCTCAGTTTTCA		1500					
1198	GACTT	CTTAGGCACCTGTGATCAACCAATCAACCGTCTGCTCAAGGGTTCGAGCGCTTCTCA		1257					
1501	GCTC	GAGGACCGAGTTTAGCCNAATCCTTCGATATCAAAAACGGATATGATTGCTCCGGGA		1560					
1258	AGTCT	GAGACCGAAACCGCTTTGCTCCAGAGATCTCTAAGCCGACATGATTTGCCCGGCT		1317					
1561	GTCA	ATATCATTTGGCGCTTGGCCTCAAAATCTAGGACCAACCC---GGACTTCCTTATGAT		1617					
1318	GTGCA	CATCTCGCTGCATGNAACCGAGAAATGTGCGCTACCATTGGCCATGTATCGAC		1377					
1618	TCAAG	AAGAGTTAACTTCACTGTATATGTACGAAACTTCAATGTCTTGTCCACATGTTAGC		1677					
1378	AATAG	CGCGCTCGAGTTTCAACATCATCTCCGGCAGCTCCATGGCGGTGCTTGCACATGAGC		1437					

QY 1678 GGAATCACTGCTCTTATCCGGTCTGCATACCCGAACTGGTCTCCAGCTGCAATCAATCC 1737
Db 1438 GGCATCGCCGCTATGCTTAAGTGGCAACACCGAGTTGGAGCCGAGCGCAATCAAGTCG 1497
QY 1738 GCATTGATGACACACAGCGGATTGTAGATCGTCAAGGAAAGCGATAAAGGATGGTAAC 1797
Db 1498 GCCATGATGAGGAGCGCCTACAAATGTGGCAATGATGGCAATGCCATCAAGGACATGGCC 1557
QY 1798 A-----AACCGCCGGTGTGGTGGATGGAGCAGGCGCATGTGAATCCGCAAAAGCG 1851
Db 1558 ACGGCGCAGCAGCCAGGCGGTTTCGAGCTTGGGTCCTGGCCACGTCGACCCCAACCGTCC 1617
QY 1852 ATAAACCCGGATTGTTTACAACTTCAACAGTGGATTACATAACTTACCTCTGCACT 1911
Db 1618 CTGACCCCTGGATTGTGTGAACACACCCGCGGACGATTATACATCTCTCTGCAGC 1677
QY 1912 CTGGAATTCACAAG 1925
Db 1678 CTGCGCTACACAG 1691

RESULT 7

US-09-770-149-59

; Sequence 59, Application US/09770149

; Patent No. US20020059663A1

; GENERAL INFORMATION:

; APPLICANT: Gorlach, Jorn

; APPLICANT: An, Yong-Qiang

; APPLICANT: Hamilton, Carol M.

; APPLICANT: Price, Jennifer L.

; APPLICANT: Raines, Tracy M.

; APPLICANT: Yu, Yang

; APPLICANT: Rameaka, Joshua G.

; APPLICANT: Page, Amy

; APPLICANT: Matthew, Abraham V.

; APPLICANT: Ledford, Brooke L.

; APPLICANT: Woessner, Jeffrey P.

; APPLICANT: Haas, William David

; APPLICANT: Garcia, Carlos A.

; APPLICANT: Kriker, Maja

; APPLICANT: Slader, Ted

; APPLICANT: Davis, Keith R.

; APPLICANT: Allen, Keith

; APPLICANT: Hoffman, Neil

; APPLICANT: Hurban, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; TITLE OF INVENTION: thaliana

; FILE REFERENCE: 2024 (PARA-013PRV)

; CURRENT APPLICATION NUMBER: US/09/770,149

; CURRENT FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: 60/178,506

; PRIOR FILING DATE: 2000-01-27

; NUMBER OF SEQ ID NOS: 999

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 59

; LENGTH: 740

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-770-149-59

Query Match 7.5%; Score 174.6; DB 9; Length 740;

Best Local Similarity 57.0%; Pred. No. 3.5e-44;

Matches 338; Conservative 0; Mismatches 252; Indels 3; Gaps 1;

QY 1165 GGATTAAGAAAGTGAATGTTCTGTTGAGAGGTCACCTCCAGAGAGAAATCCGAGCC 1224
Db 131 GCGGATTGCGGAGACAGATTGTGTACCTGGGAATTTGAATTCATCATTTGGTTGAAGCC 190
QY 1225 AAAATGTTGATTTCTGTGATCGCGGAGTCAATGGAAGATCGGAGAAAGAGAGCGGTTAAA 1284
Db 191 AAAATGTTGTTCTGTGACAGAGGAGGCAACGCAAGAGTTTGAGAAAGAGTGCAGTCAAG 250

QY 1285 GAAGCTGGAGGAGTTGCAATGATCTTAGCCAAATACAGAGATCAACCAAGAGAGATTCT 1344
Db 251 CTAGCCGGTGGTCTGCTGATGATTTCTGGGAAACACAGCTGAAAGCGGTGAAGAAATTAACC 310
QY 1345 ATTGACGTTTCACTCTTACCAGCTACATTGATTGGTTTACACTGAGTCAGTCTTCTGAAG 1404
Db 311 GCGGATTTCGATCTCGTCCGCGCAATGGTTGGAGCTTAAGCTGGAGATCAATCCGC 370
QY 1405 GCTTATTATTAATGCCAGGTGAAACCAAGGCGCGGATAATTTTGGTGGTACGGTGAAT 1464
Db 371 GACTACATCAAAACATCAGACTCTCCACTGCAAAAATCAGTTTCTTAGSCACTTTGATC 430
QY 1465 GGGAGGTACGA---GCACCGGAGGTGGCTCAGTTTTTTCAGCTCGAGGACCGAGTTTACGC 1521
Db 431 GGACCACTCTCTCTCTCCAGAGTCGCGCTTTTCTCCAGCGCTGGACCGAATCACTTG 490
QY 1522 ATCTCTTCGATATAAACCGGATATGNTTCTCCGGAGTCAATATCATTTGGCGGTTGG 1581
Db 491 ACACCGGTTATTCTTAAACCGGAGCTGATTGCTCTCGGAGTCAACATTTTAGCCGTTGG 550
QY 1582 CCTCAAAATCTAGGACCAACCGGACTTCTCTTATGATTCAAGAAAGATTAACTTCACTGTA 1641
Db 551 ACGGGATGGTTGGTCTCTACCGATTAGATATCGATCCAGAGCGGTTCAATTCACATC 610
QY 1642 ATGTCAAGAACTTCAATGCTTTGTCACATGTTTAGCGGAATCACTGCTCTTATCCGGTCT 1701
Db 611 ATCTCCGTTACATCGATGCTGTCGCCACACGTTAGTGGAGCTCGCGCTCTCTCCGTAAA 670
QY 1702 GCATACCCGAACTGCTCTCCAGCTGCATCAAAATCCGATTGATGACACAGC 1754
Db 671 GCTCATCCGATTGGTCACTCGAGCAATCAAAATCCGCTTGTAAACCCCGC 723

RESULT 8

US-10-260-238-5684/c

; Sequence 5684, Application US/10260238

; Publication No. US20040016025A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, Paul R.

; APPLICANT: Moughamer, Todd G.

; APPLICANT: Briggs, Steven P.

; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane

; APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyaki

; APPLICANT: Kreps, Joel

; APPLICANT: Provart, Nicholas

; APPLICANT: Ricke, Darrell

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION

; FILE REFERENCE: 6011-NP

; CURRENT APPLICATION NUMBER: US/10/260,238

; CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/325,448

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/370,620

; PRIOR FILING DATE: 2002-04-04

; NUMBER OF SEQ ID NOS: 6077

; SEQ ID NO 5684

; LENGTH: 606

; TYPE: DNA

; ORGANISM: Zea mays

US-10-260-238-5684

Query Match 5.6%; Score 153.2; DB 12; Length 606;

Best Local Similarity 55.4%; Pred. No. 2e-37;

Matches 296; Conservative 0; Mismatches 238; Indels 0; Gaps 0;

QY 1221 AGGCAAAATGTTGATTGTTGATCGCGAGTCAATGGAAGATCGGAGAAAGAGAGCGGT 1280
Db 537 AGGCAAGATGTTGTTCTGCGACCGCGCGGCGAGCAGCGCGCGGTGCGCAAGGAATGTCGT 478

QY 1281 TAAAGAGCTGGAGGAGTTCATATGATCTTAGCCATACAGAGATCAACCAAGAGAGA 1340
DB 477 CAAGGAAGCTGGTGGCGCGCATGCTTCTTAAACGGGAGAGCCCAACGGCGAGGACT 418
QY 1341 TTCTATTGAGCTTCATCTTTACAGCTACATGATTTGTTTACACTGAGTCACTCTCTCT 1400
DB 417 GATTGGAGAGCGCCACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 358
QY 1401 GAAGGCTTATGTTAATCCACCGGTGAACCAAGAGCGCGGATTAATTTTGGTGTACGT 1460
DB 357 CAAGCGTATGCCGGAACGCTCCAGCCGATGGGAGACATCAGCTTCGGGGGACGCT 298
QY 1461 GATTGGAGGTACAGACACCGGAGGTGCTCATGTTTTCAGTCTGAGGACCGAGTTTACG 1520
DB 297 CGTGGCGGTCAAGCGCGCACCGGTGGTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 238
QY 1521 CAATCTTCGATTAACCAACCGGATGATGTTCTCGGGAGTCAATATCATTTGCGGCTTG 1580
DB 237 CGTCCCGGAGATTCGAAGCGGACTTCATCGCACCGGGGTCAACATCTCGCCGCTG 178
QY 1581 GCCTCAAAATCTAGGACCAACCGGACTTCCTTATGATTCGAAGAGATTAATCTCACTGT 1640
DB 177 GACGGGCGCACGGTCCGACCGGCTGGAGGGGATACCCGCGGACGGAGTTCAACAT 118
QY 1641 AATCTCAGGAATCTCAATGTTTTCACATGTTAGCGGAATCACTGCTTTATCCGCTC 1700
DB 117 CTTGTCGGAGCTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 58
QY 1701 TGATACCGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1754
DB 57 CGCGACCCCGGGTGGTGGCGGGGGATACGCTCGCGCTGATGACTACGGC 4

RESULT 9

US-10-260-238-703
; Sequence 703, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 703
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (354)..(354)
; OTHER INFORMATION: n = any nucleotide

Query Match 6.1%; Score 141.4; DB 12; Length 1467;
Best Local Similarity 51.7%; Pred. No. 2.1e-33;

Matches 378; Conservative 0; Mismatches 341; Indels 12; Gaps 2;
QY 395 GGTGTATGCTCTAAATCTCGGTTTGGTCAAGGCACATTTATTCGCGTCTGTATCTGGA 444
DB 367 GGCCTGTGGGAGACGGCCAGCTACGGCGACGGCGTCTATCGTCCGCGTGGTGACACCGGG 426
QY 445 GTTTGGCTGAAAGTCTCTAGCTTTGACGATACCGGAATGCTTCTGATTTCCACGGAATGG 504
DB 427 GTTTGGCCGAGAGCGGAGCTACCGGACGACGGGCTCCGCGGCTCCGCGCGGTGG 486
QY 505 AAAGGGATTTCCCAAGAGAGAAAGTTTCA---GTTCTTCGAGCTGTAAACCGGAAGCTA 561
DB 487 AAGGGCTACTCGGAGTCCGGCACCGGTTTCAAGCGGCCAAGGGGTGCAACCGGAAGCTC 546
QY 562 ATCGGTGTGATTTCTTCATCAGAGGACACCGTGTCTGATTAATTCACAGAGGAATACCA 621
DB 547 ATCGGGGCGCAAGTTTCAAGCGGCGCTCGCCCGCGCT-----GGGAGGCGG 597
QY 622 AACATGCTCTGATATCATTTCCGCAAGAGATTTCAACGGGACACGGGACTCACACCGCC 681
DB 598 AACATCAGATCGCGGTGAATTCGCGCGGACACGGACGGCCACGGACGCAACCTCG 657
QY 682 TCAACAGTTGTTGATCTCTGTTTCGATGGCGAATGTTCTTGGCAATGAGCTGGTGTG 741
DB 658 TCGACGGCGCGGCTCGCCGCTCCGCGGGCGTCTTCTTCGCTACGCTCCCGGGGTG 717
QY 742 GCTGTGGGATGGTCTCTGGAGCTCACATTCAGTCTATAAAGTCTGTGTTCAATGGT 801
DB 718 GCACGGGCGATGGCGCCACGGCGCAGGCTGGCGCTGTACAGGTCTCTGTTTCGACGAGGC 777
QY 802 TGTTACAGCTGTGATCTTAGAGCTATAGATGTAGGATTAAGATAAAGTCTGATGTT 861
DB 778 GGGTACAGCGGACATCGTCCGCGCCATAGCAAGCCATCGCAGACGGGCTCGACGTG 837
QY 862 CTTTCGCTTTCCTTGGCGGTTTCCCTATTCTTGTATGATGACACAATCGCATTTGA 921
DB 838 CTCTCCATCTCGTGGGCTTCAACACCGCCCGTCCACCGACCCCTGTGCGCATCGGC 897
QY 922 ACATTTCCGAGCATGGAACGGGTTATCTGTAATCTGTGCACTGTGTAAACAGGTCCA 981
DB 898 TCGTTCCGGCAATGCAGCAGCGCATCTTCGTGTCCACCTCAGCGCGCAACGCGGCCG 957
QY 982 ATCGAAAGCTCTGTTGCAACACAGCTCTTGGGTCTCAACATTTGGCGAGGACCGTT 1041
DB 958 GGCCTCTCCGCTCTCCACACGGCGCGCTGGGCACTGACCGTCCGCGCTGGCACCGTG 1017
QY 1042 GATCGAAGTTTCGCGTGTGTCAGATTAGCCCAACGGAAGCTTCTATGTAGAGTCA 1101
DB 1018 GACAGGGAGTTTCTCCGGCATCTCGAGCTCGGATGCGACACCGTATCGGCGAATCG 1077
QY 1102 TTGTATCCGGG 1112
DB 1078 CTGTACGCGG 1088

RESULT 10

US-10-260-238-1133
; Sequence 1133, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP

```

; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 1133
; LENGTH: 2334
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (498)..(498)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N region
; LOCATION: (548)..(548)
; OTHER INFORMATION: n = any nucleotide
; NAME/KEY: N region
; LOCATION: (551)..(551)
; OTHER INFORMATION: n = any nucleotide
; US-10-260-238-1133

Query Match      5.2%; Score 121.2; DB 12; Length 2334;
Best Local Similarity 46.8%; Pred. No. 8.1e-27;
Matches 593; Conservative 0; Mismatches 653; Indels 20; Gaps 6;

QY 661 GGACACGGGACTCACACCGGCTCAACAGTTGGTGGATCTCTGTTTC---GATGGCGAAT 717
DB 652 GGCACGGGACACACCGGCTCCAGCGCGCGGAACTTCGTCCCGCGCGCGCTCG 711
QY 718 GTTCTTTGGCAATGGAGTGTGTGGCTCGTGGGATGCTCTTGAGCTCACATTCGAGTC 777
DB 712 GACCGGGCTGGCGCGGGAACGGCGCGGATCGCGCGCGGCGCACGTCGCGCATG 771
QY 778 TATAAGTCTGTGGTTCAATGTTTACAGCTCTGACATCTAGCAGCTATAGATGTA 837
DB 772 TACAAGGTGTGCAACGGCTCAGGCTCGATGACAGCGCGTGTAGCCGGGTTTCGACGAG 831
QY 838 GCGATTCAAGATAAGTCGATGTTCTTCCTTTCCCTTGGCGGTTTCCTATT---OCT 894
DB 832 GCCATGAAGACGGAGTCGACGCTCTCTCGGTCTCGCTCGGAGGTGTCGAGTCCCCC 891
QY 895 TTGTATGATGACACAAATCGCCATTGGAAACATTCGAGCCATGGAACGCGGTATATCTGTA 954
DB 892 TTGATGAGACCCCATCGCATCGCGCTTCAGCGAGTGGCGGAGGCATCACCGTG 951
QY 955 ATCTGTGACGTGTGTAACACCGTTCGAATCGAAAGCTCTGTGCAAAACACAGCTCTTGG 1014
DB 952 GTGTGCGCGCGGCAATGGCGGCGCGCGCGCGCTCGACGGTCTCGAAGCAGCGCGCGTGG 1011
QY 1015 GTCTCACCATTTGCGGACGACGCTTGATCGAGATTTCCGCTGTGTCAGATAGCC 1074
DB 1012 CTGCTCACGTTGCGCGGAGCTGGTCTCGCGGAGCTTCTCCACCACCGTGTGCTCGGC 1071
QY 1075 AACGGAAGCTCTCTATGAGAGTCAATGTATCCGGGAAAGGTATAAGAAATCCGGG 1134
DB 1072 AACGGGAGCTCTGACGCGCGAGCGCT--CGCCAGCAGCCCACTCTCGACCTCGTA 1130
QY 1135 AGAGAGGTGTAGGTGATTACGTACAGGAGGAGATAAGGAAGTAGTGTCTGTTTGA 1194
DB 1131 CTACCCACTCCACTTCTCGGAGAAGCAACCGAAATGTAACGAGCTCGCGGAATCGTCG 1190
QY 1195 GGTCACTTCAAGAGAGAAATCCGAGGCAAAATGTTGATTCGCGAGTCAAT 1254
DB 1191 AGATGGGCTCGCGCGCACCTCTGTGTCTGCCAGTCCGATCCGCTCGAGGACGATCCGT 1250
QY 1255 GGAAGATCGGGAAGGAGAGCGGTTAAGAGCTGGAGAGTTGCA---ATGATCTT 1310
; TITLE OF INVENTION: Arabidopsis thaliana

1251 CGTAGGGCCATGATGGCCACCGCGCGCGCGGTGCTGATCAACACCGAGAGCA 1310
1311 AGCAATACAGAGATCAACCAAGAAAGATTTATTTGAGTTCACTCTTTACAGCTAC 1370
1311 AGCTACAGACCGCTCTCGAGGACTACGCGCGGCGCATGGTGCAGGTGACCGTCCGCG 1370
1371 ATTGATTGTTACATGAGTCACTCTTCTGAGGCTTATGTTAATGCCACGGTGAACC 1430
1371 TGGCCCAACATCACGAGTACGCTAGTCTGCTCATCATCAGCGCGGCTGCAAAAC 1430
1431 AAAGGCGCGGATAATTTTGGTGTAGTGGAGGTCAAGAGCAACCGAGGTGGC 1490
1431 CAACGCCACCGTGTGTGCAAAACAGTTGCTCAGCGTCCACCGCGCGCCACCGTGC 1490
1491 TCAGTTTTCAGTCTCGAGGACCGAGTTTAGCCAACTCTTCGATACTAAAAACCGATAT 1550
1491 CTCTTCTCTCGCGGACCGAGAGTTCGCGCGCGGCTCAAGCGGACGCTCT 1550
1551 TGCTCGGAGTCAATATATTCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCC 1610
1551 GGCGCGGGCTCAACATCTCTCGCGCATGGCGCGGCTCCAGCACCGCGG----- 1604
1611 TTATGATTCAAGAGAGTTAACTTCACTGTAATGTCAGGAACCTTCAATGTTTCCACA 1670
1605 CGCGCGGAGGCGCGGCTCTTCAAGTTCATCCGAGCTCCATCGGACCGCCCGCGCA 1664
1671 TGTAGCGGAATCACTCTCTTATCCGCTCTGATACCCGAACTGCTCTCCAGCTCAAT 1730
1665 CGGAGCGGCTGGCGCGCTCTCAAGACCGCCACCGGACTGTTGCGCGCGCCAT 1724
1731 CAATCCGCTTGTATGACAAACAGCGGATTTGTACGATCGTCAAGGGAACCGATAAGGA 1790
1725 CAAGTCAGCCATCTGACGACGCTCCGACGCGCTCGACGCGCGGCAACCGGATCTTGA 1784
1791 TG---GTAACAAACCGCGCTGTTGCTGATGAGGAGGCACTGTAATCGCAAAA 1847
1785 CGAACACACGAGAGGCGGCTCTCAACCGCGCGGCGCACTCAACCGCGCGAG 1844
1848 GGCGATAAACCGGAGTTGTTTACACATTTCAACAGTGGATTACATAACTTTACCTCTG 1907
1845 GGCGCGGACCGAGGCTGTAGTACGACATCGCGCTGCTGCTACCGCGCTACATCTG 1904
1908 CACTCT 1913
1905 CGCGCT 1910

RESULT 11
US-09-770-961-228
; Sequence 228, Application US/09770961
; Publication No. US20030115639A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurlan, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis thaliana
```

FILE REFERENCE: 2026 (PARA-015PRV)
CURRENT APPLICATION NUMBER: US/09/770,961
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,466
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 228
LENGTH: 510
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-961-228

Query Match 4.4%; Score 103.2; DB 11; Length 510;
Best Local Similarity 54.5%; Pred. No. 1.5e-21;
Matches 207; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 200 CTTCTTCTCGACTTCTCTACTCCTTGTGCTCTGCGATTGAAGGATTGCTCGCTCAGTTGA 259
Db 128 CTTCTTCTCGACTTCTCTACTCCTTGTGCTCTGCGATTGAAGGATTGCTCGCTCAGTTGA 187
QY 260 CTGAATCAGAACCGGAGATCTGAGATATTCACCTGAAGTTGTTGCACTGAGTGAACCTGACC 319
Db 188 CTTACAGAGCGTAGTCACTCTTACACCATCTCTCATGTCATCTCTGTTATCCCGGAAC 247
QY 320 ATGTTCTTCAAGTTCAAAACCACTTACTCTTACAAAGTTCTTTGGGACTCGACGGTTTGGAA 379
Db 248 AAGTCCGGTCACTTGCACACCACTCGTTCCTCCCTGAGTTCTTTGGTCTTTAGTCCACCGACA 307
QY 380 ACTCGGTGATGTCTAAATCTCGTTTGTGTCAGGACAAATATCGGGTGTCTTGATA 439
Db 308 AAGCGGTCTACTCGAAGAGTCTGATTTCCGGTCTGATCTAGTTATCGGAGTTATCGATA 367
QY 440 CTGAGTTTGGCTTGAAGTCTTACGTTTTCAGCATCCGGAATGCCCTTCGATCCACGGA 499
Db 368 CTGTTGTTGGCCGGAAGACCTAGCTTTGATGACCTGTCTGTTGCTCTGTTTCCCATTA 427
QY 500 AATGGAAGGATTTGGCAAGAGGAGAAAGTTTTCAGTTCTTTCAGCTGTAAACCGGAAGC 559
Db 428 AATGGAAGGCAATGATCGTCCCAAGATTTCCGGAGTCTGCTTGAACGGTAAC 487
QY 560 TAATCGGTGCTAGATTTCTTC 579
Db 488 TCGTCCGAGCTAGATTTCTTC 507

RESULT 12

US-09-938-842A-338
Sequence 338, Application US/09938842A
Publication No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 338
LENGTH: 2958
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-338

Query Match 4.3%; Score 100.2; DB 10; Length 2958;
Best Local Similarity 56.5%; Pred. No. 4.7e-20;
Matches 186; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 1425 GAAACCAAGCGCGGATAATTTTGGTGTACGGTGAATGGGAGGTCCAGGACCGGA 1484
Db 1701 GAAGCGGATTGCACACATTACTCTTCAAGGACAGATTGGGACTAAACACGCGCTGT 1760
QY 1485 GGTGCTCAGTTTTCAGTCCGAGGACCGAGTTTAGCCAAATCTTCGATACATAAAACCGGA 1544
Db 1761 AATGGCTTCATTTTCTTCAAAAGGTCGAAGCATAGTGGCTCTCAGATTCTGAAGCTGA 1820
QY 1545 TATGATTGCTCCGGAGTCAATATCATTTGGCGTTGGCTCAAAATCTAGACCAACCGG 1604
Db 1821 CATAACTGCTCTGCTGTGAGTGTATCGCTGCTACACTGGAGCAGTCTCTCCAACAAA 1880
QY 1605 ACTTCTTATGATTTCAAGAAGAGTTAACTTCACTGTAATGTGAGAACTTCAATGCTCTTG 1664
Db 1881 CGAACAGTTTGAATCTCGACGACTTCTGTTCAATGCTAATTCAGGAACCTTATGCTCTTG 1940
QY 1665 TCCACATGTTAGCGGAATCACTGCTTCTTATCCGGTCTGCATACCCGAACTGGTCTCCAGC 1724
Db 1941 TCTTCATATCTCGCAATTCGGGCTTCTCAAAACTCGTTATCTCTTGGAGCCCGC 2000
QY 1725 TGCATCAAAATCCGCAATTGATGACACAG 1753
Db 2001 AGCTATCGGCTCTGCCATCATGACTACCG 2029

RESULT 13

US-09-938-842A-338
Sequence 338, Application US/09938842A
Publication No. US20040009476A9
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 338
LENGTH: 2958
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-338

Query Match 4.3%; Score 100.2; DB 12; Length 2958;
Best Local Similarity 56.5%; Pred. No. 4.7e-20;
Matches 186; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 1425 GAAACCAAGCGCGGATAATTTTGGTGTACGGTGAATGGGAGGTCCAGGACCGGA 1484
Db 1701 GAAGCGGATTGCACACATTACTCTTCAAGGACAGATTGGGACTAAACACGCGCTGT 1760
QY 1485 GGTGCTCAGTTTTCAGTCCGAGGACCGAGTTTAGCCAAATCTTCGATACATAAAACCGGA 1544
Db 1761 AATGGCTTCATTTTCTTCAAAAGGTCGAAGCATAGTGGCTCTCAGATTCTGAAGCTGA 1820
QY 1545 TATGATTGCTCCGGAGTCAATATCATTTGGCGTTGGCTCAAAATCTAGACCAACCGG 1604
Db 1821 CATAACTGCTCTGCTGTGAGTGTATCGCTGCTACACTGGAGCAGTCTCTCCAACAAA 1880
QY 1605 ACTTCTTATGATTTCAAGAAGAGTTAACTTCACTGTAATGTGAGAACTTCAATGCTCTTG 1664

Db 1881 CGAACAGTTGATCTCTGACGACTTCTGTTCAATGTTATTTAGGAACCTCTATGCTTG 1940
Qy 1665 TCACATGTTAGCGGAATCACTGCTTTATCCGGTCTGCATACCCGAACTGGTCTCCAGC 1724
Db 1941 TCCTCATATCTTGGCAATTCGGGGTCTTCTCAAAATCGTTATCCTTCTTGGAGCCCGC 2000
Qy 1725 TGCATCAAAATCCGCAATGATGACACAG 1753
Db 2001 AGCTATCCGCTCTGCCATCATGACTACCG 2029

RESULT 14
US-10-260-238-3023
; Sequence 3023, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 3023
; LENGTH: 849
; TYPE: DNA
; ORGANISM: *Triticum aestivum*
; FEATURE:
; NAME/KEY: N region
; LOCATION: (814)..(814)
; OTHER INFORMATION: n = any nucleotide

US-10-260-238-3023
Query Match 4.2%; Score 98.4; DB 12; Length 849;
Best Local Similarity 51.0%; Pred. No. 7.3e-20;
Matches 358; Conservative 0; Mismatches 311; Indels 33; Gaps 4;
Qy 1496 TTTGAGTTCGAGGACCGAGTTAGCCCAATCTTCGATACATAAAACCGATATGATGCTC 1555
Db 1 TCTTTCCGGTGGTCCAGCGTCTCGCCCTCGCATCTTAAGCCGGATATATTGGCAC 60
Qy 1556 CGGAGTCAATATCATTTGCGGCTTGGCTTGAATCTAAATCTAGACCAACCGGACTTCTTTATG 1615
Db 61 CAGGCCTCAATATCTCGCCCGTGGCCGCAAGCAGC----- 101
Qy 1616 ATTCAGAGAGATTAACTTCACTTAATGTCAGGAACCTCAATGCTTTGTCACATGTTA 1675
Db 102 --TCTGGAAGGGGCGCTTTCAACATCATATCAGGAACATCATGGGACGCCACAGTCA 159
Qy 1676 GCGGAATCACTGCTTTATCCGTTCTGCATACCCGAACTGGTCTCCAGCTGCAATCAAAAT 1735
Db 160 GTGGTGCTTGCACTTATCGAAGCATCCATCTCTGACTGGTCCGACGCCCATCAAGT 219
Qy 1736 CCGCATTTGATGACACACGCGGATTGTACGATCTGTCAGGAAAGCGATAAAGATG--- 1792
Db 220 CTGCATCTCAACGACGTCGACATCGTTATAGCACTGGTGGCTCAATCTTGGACGAGC 279

Qy 1793 GTAACACACGCGGTGTTGTTTCGATTGAGCAGGCGCATGTGAATCCGCAAAAGCGCA 1852
Db 280 AACATAGGAAGCGCGCGTGTACGACACAGCGCGCGCCATGTGAACCCGACGAGCGC 339
Qy 1853 TAAACCCGCGATTGTTTACAAATTCACACAGTGGATTACATAAATTTACTCTGTGCACTC 1912
Db 340 CAGATCTTGGCTTGGTGTACGACCTAGTCTCACTAGTACCGCGGCTACATCTGTGTCG 399
Qy 1913 TTGGATTCAAGAATCAGATATTTTAGCAATCACTATAAGAACGTCGAGCTGCAATGGAA 1972
Db 400 TCCTAGGTGACAG--CGGCCTAGCAACCATCGTCGCAACTCAAGCTTCAACCTGTGCAA 456
Qy 1973 TATTGCGGAAAAACCCCGGTTTGTCTCAATTAATCCGTCGATAGCCGCTGATTTCAAAC 2032
Db 457 AGCTGCCCAAGGTCCACGCGCTGAGCTTAATTAATCCGCGATTAACCGTGTACGCGACA- 515
Qy 2033 GTGGCAAGACTACGAGATGATCAAGGCGTGTCACTAACTGGGAGTCTCTTAACCTCGA 2092
Db 516 -----TCGACGCCATTACAGTAAGCGGACTGTGACAAACGTTGGCCAGCAACTCGA 570
Qy 2093 TATACTCAGTGAATGTCAAGGCTCCAGAGGGGATCAAAAGTTATTTGTCAATCTTAAGAGAC 2152
Db 571 CATTACAGTGAAGGTGACGCGCAAGATCACTGACAGTGGGTCTCTCCCGGAGACGC 630
Qy 2153 TTGTGTTCAACACAGTGGATCAGCGTGGCTATAGATAT 2194
Db 631 TGGCATTCTCCAAAACCGGAGAGAAAGACCTTTCAGCGTGT 672

RESULT 15
US-10-260-238-3022
; Sequence 3022, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 3022
; LENGTH: 828
; TYPE: DNA
; ORGANISM: *Triticum aestivum*
US-10-260-238-3022

Query Match 4.0%; Score 92.4; DB 12; Length 828;
Best Local Similarity 51.2%; Pred. No. 5.9e-18;
Matches 297; Conservative 0; Mismatches 271; Indels 12; Gaps 3;
Qy 1618 TCAAGAAGATTAACTTCACTGTATGTCAAGAACTTCAATGCTTTGTCCACATGTTAGC 1677
Db 64 TCTGGAAGGGCACATTTCAACATCATATCAGGAACATCATGGGACGCCACAGCTCAGT 123
Qy 1678 GGAATCACTGCTTTATCCGGTCTGTCATACCCGAACTGGTCTCCAGCTGCAATCAAAATCC 1737
Db 124 GGTGTCGTGGCTTATCAAAAGCATCCATCTGACTGGTGGCCAGCGCCATCAAAATCT 183

QY	1738	GCATTTGATGACAAACAGCGGATTTGTACGATCTGTCAGGGAAACGGATAAAGGATG---GT	1794
Db	184	GCCATCCTAAACGAGCTCGGACATCGTTTAATAGCACGTGGTGCTCAATCTTTGGACGAGCAA	243
QY	1795	AACAAACACAGCCGGTGTGTTTGCATTTGGAGCAGGGCATGTGAATCCGCAAAAGCGATA	1854
Db	244	CATAGGAAGCCGGCTGTATGACACAGGCGCCGCCATGTGAAACCGACCAAGACCGCA	303
QY	1855	AACCCGGATGGTTTAAACATTCACACAGTGGATTACATAACTTACCTCTCGACTCTT	1914
Db	304	GATCCTGGCTTGGTGTACGACCTCAGCGTCACCGACTACGCCGGCTACATCTGCTGGCTC	363
QY	1915	GGATTCACAGATCAGATAITTTAGCAATCACATCAATAGAACGTGAGTGCATATGGAATA	1974
Db	364	CTCGGTGACAGC---GGCCTAGCAACCATCGTGGCAACTCAAGTTGACCTGTGGGAAG	420
QY	1975	TTGCGGAAAAACCCGGGTTTTAGTCTCAATTTACCCGTCGATAGCCGTGATTTTCAAACGT	2034
Db	421	CTGCCAAGGTCCATGCTGTGCAGCTTAACTACCGACGATAACCGGTGTGACGCGACATC-	479
QY	2035	GGCAAGACTACGGAGATGATCACAAAGCGGTGTCACTAACGTTGGGAGTCCCTAATCGATA	2094
Db	480	-----GACGCCATTCAACAGTGAACGCGACTGTGACGAAACGTTGGGCGACAAAACCTCGACG	534
QY	2095	TACTCAGTGAATGTCAAGGGTCCAGAGGGGATCAAAGTTATTGTCTAACTCCTAAGAGACTT	2154
Db	535	TTACGCGGGAAGGTGGACGGCCCAAGATCACTGACGTGGGTGTCTCCCGGAGACGCTG	594
QY	2155	GTGTTTCAAAACACGTGGATTCAGACCGCTGAGCTATAGATAT	2194
Db	595	GGATTTCTCCAAAACCGGAGAGAAGAAGACCTCTCAGCGTGT	634

Search completed: February 3, 2004, 02:18:36
Job time : 630 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 21:36:20 ; Search time 3282 Seconds
(without alignments)
17239.738 Million cell updates/sec

Title: US-09-806-767-1
Perfect score: 2328
Sequence: 1 atggaacccaacactttttt.....taacctgaagactaactga 2328

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpi:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	574.8	24.7	832	28	BH480160
c 2	475	20.4	696	28	BH660501
c 3	472.4	20.3	676	28	BH998381
c 4	461	19.8	670	28	BH464077

c 5	438.2	18.8	635	28	BH599290
c 6	426.8	18.3	627	28	BH597572
c 7	424	18.2	640	28	BH442444
c 8	409.2	17.6	803	28	BH601308
c 9	356.2	15.3	527	28	BH714875
c 10	326.8	14.0	767	28	BH793276
c 11	296.8	12.7	568	10	BF052773
c 12	273	11.7	772	28	BH601319
c 13	270.2	11.6	789	29	CC340521
c 14	252	10.8	750	28	AZ046831
c 15	233.8	10.0	713	28	BH720365
c 16	228	9.8	412	9	AV420480
c 17	223.6	9.6	719	13	BQ870650
c 18	219.4	9.4	848	14	CD575777
c 19	217.6	9.3	761	28	BH442432
c 20	215.4	9.3	1572	11	AY107161
c 21	207.6	8.9	738	13	BQ996378
c 22	206.6	8.9	697	13	BU895782
c 23	205	8.8	790	28	BH590573
c 24	203.6	8.7	762	28	BH649984
c 25	201.8	8.7	863	10	BF278122
c 26	199	8.5	735	13	BQ893301
c 27	198	8.5	800	14	CB892659
c 28	197.4	8.5	864	28	BH648619
c 29	196.8	8.5	766	28	BH725669
c 30	195.6	8.4	703	13	BQ999661
c 31	193.8	8.3	719	28	BH668947
c 32	190.2	8.2	662	28	BH434517
c 33	190.2	8.2	734	13	BQ866897
c 34	189	8.1	926	10	BG45439
c 35	188.8	8.1	610	13	BH883192
c 36	188.8	8.1	639	9	AW720622
c 37	188.4	8.1	698	28	BH994888
c 38	187.8	8.1	716	28	BH659528
c 39	186.8	8.0	526	9	AV527216
c 40	186.4	8.0	640	10	BF005920
c 41	184.8	7.9	740	13	BQ867168
c 42	184.4	7.9	558	28	BH714885
c 43	183	7.9	709	13	BQ866922
c 44	183	7.9	818	12	BI405859
c 45	182.8	7.9	639	9	A1730592

ALIGNMENTS

RESULT 1
BH480160/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BH480160 832 bp DNA linear GSS 13-DEC-2001
BOHME48TF BOHM Brassica oleracea genomic clone BOHME48, genomic
survey sequence.

BH480160 GI:17698264

GSS.

Brassica oleracea

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosid II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 832)

Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished

Other_GSSs: BOHME48TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

```

FEATURES
  source
    Location/Qualifiers
      1..832
        /organism="Brassica oleracea"
        /mol_type="genomic DNA"
        /strain="TO100DH3"
        /db_xref="taxon:3712"
        /clone="BOHME48"
        /notes="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
        genomic DNA inserted into PHOS1 using BstXI linkers"
      177 a 209 c 247 g 199 t

BASE COUNT 177 a 209 c 247 g 199 t
ORIGIN
  Query Match 24.7%; Score 574.8; DB 28; Length 832;
  Best Local Similarity 81.9%; Pred. No. 7.2e-129;
  Matches 663; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
  QY 248 CTGCTGATGCTGATCAAGAGCGAGATACAGATGATTCACCTGAAGTTGTTGAG 307
  Db 831 CAGCTCAGTTAACCAGATCAGAGCCAAACACACTCAAGACTTTACCTGAAGTTGTTG 772
  QY 308 TGAGACTGACCATGTTCTTCAGGTTCAACACCACTTACTCTTACAGTTCTTTGGGACTCG 367
  Db 771 TAAGACCTGACCATGTTCTTCAAGTCCAAACCACTTACTCTTACAGTTCTTTAGGACTCA 712
  QY 368 ACGGTTTGGAAACTCCGGTGTATGTTCTAAATCTCGGTTTGTCTAAGGCACAAATTATCG 427
  Db 711 ACGGTCAGACCCCTCCAGCGTATGTTCTAAATCCCGGTCGCGCCAGGCACGATCATCG 552
  QY 428 GCGTGTGTGATATGGAGTTTGGCTGAAAGTCTTAGCTTTGACGATACCGAAATGCTTT 487
  Db 651 GCGTTCTCGACACAGAGTCTGCGCCGAAAGCCCTAGCTTCGACGACACAGGATGCTTT 592
  QY 488 CGATTCCACGGAATGGAAGGATTTGCCAAGAGGAGAAAGTTTCAGTTCTTCGAGCT 547
  Db 591 CAGTCCCAAGCAATGGAAGGTTTGGCCAAGAGGAGAGCTTCACTTCTCAAACT 532
  QY 548 GTAACCGGAGCTAATCCGGTGTAGATTCTTATCAGAGGACACCGTGTCTAATTCAC 607
  Db 531 GTAAACAGAACTAATCCGGCGCTAGATTCTTATCAGAGGCAACCGGTAGCACTTC 472
  QY 608 CAGAGGAATCACCACCAATCCCTCGTGAATACATTTCCGCAAGAGATTCAACGGGACAG 667
  Db 471 CTTTGGACTCACCACCAATCCCTCGGATACATATCGGCACGAGACTCCACGGGSCAG 412
  QY 668 GGACTACACGGCTCAACAGTTGGTGGATCTCTGTTTTCGATGGGGAATGTTCTTGGCA 727
  Db 411 GGACCCACACCGCTCAACTGCTGGCGGCTCTCTGTTTCGACGGCGAGCGTTCTCGTA 352
  QY 728 ATGAGCTGTTGGCTCGTGGGATGCTCTCTGAGCTCACATTGCACTTATAAGTCT 787
  Db 351 ACGAGCCGGGTAGCCCGCGGATGCGCCAGAGCTCACGTGGCGGTTTCAAAAGTCT 292
  QY 788 GTTGGTTCAATGTTGTTACAGCTCTGACATTTCTAGCAGCTATAGATGTAGCGATTCAAG 847
  Db 291 GCTGTTCAACGCTGTTACAGCTCAGACATTTCTGGCGGGATAGAGTAGGATACAAG 232
  QY 848 ATAAAGTCGATGTTCTTCCCTTCCCTTGGCGGTTTCCCTATCTCTTGTATGATGACA 907
  Db 231 ACAAGTCGACGTTCTATCCCTCTCCCTAGCGGTTTCCCTATCCCTTGTATGATGACA 172
  QY 908 CAATCGCCATTGGACATTTCCGAGCCATGGAACGCGGTATATCTGTAACTGTGCACTG 967
  Db 171 CTATCGCCATTGGTACCTTCCGCGCCACGGAATGGAATCTCCGTGTCTGCGCGGGTG 112
  QY 968 GTAAACACGGTCCCAATCGAAAGTCTGTGTGCAAAACACAGCTCTCTGGGTCTCAACATTG 1027
  Db 111 GCAACACGGCCCGCTCGCTAGCTCGGTGCGGAACACAGCTCTCTGGGTCTCAACCTCG 52
  QY 1028 GCGCAGGACGCTTGATCGAAGATTTCCCG 1057
  Db 51 GAGCAGGAACGCTCGATAGAAAATTTCCCG 22

```

```

RESULT 2
BH660501
LOCUS BOHWS71TF BO_2_3 KB Brassica oleracea genomic clone BOHWS71,
        696 bp DNA linear GSS 19-FEB-2002
        genomic survey sequence.
ACCESSION BH660501
VERSION BH660501.1 GI:18719169
SOURCE GSS.
ORGANISM Brassica oleracea
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids
          ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 696)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Chris Town
        TIGR
        9712 Medical Center Drive, Rockville, MD 20850, USA.
        Tel: 301-838-3523
        Fax: 301-838-0208
        Email: ctown@tigr.org
        DNA is from a doubled haploid provided by Tom Osborn.
        Seq primer: TF
        Class: sheared ends.
FEATURES
  source
    1..696
      /organism="Brassica oleracea"
      /mol_type="genomic DNA"
      /strain="TO100DH3"
      /db_xref="taxon:3712"
      /clone="BOHWS71"
      /clone_lib="BO_2_3 KB"
      /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
      genomic DNA inserted into PHOS1 using BstXI linkers"
    163 a 166 c 220 g 147 t

BASE COUNT 163 a 166 c 220 g 147 t
ORIGIN
  Query Match 20.4%; Score 475; DB 28; Length 696;
  Best Local Similarity 80.5%; Pred. No. 1.3e-104;
  Matches 556; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
  QY 685 ACAGTTGGTGGATCCTCTGTTTCGATGGCGAATCTTCTTGCATGAGCTGGTGTGGCT 744
  Db 6 ACTGCTGGCGCTCCTCTGTTTCGACGGCGAGCGTTCGTTAACGGAGCCGGGTAGCC 65
  QY 745 CGTGGGATGGCTCCTGGAGCTCACATTGCGAGTCTATAAAGTCTGTGGTTCAATGGTTGT 804
  Db 66 CGCGGATGGCGCCAGAGCTCAGTGGCGGTTTACAAAGTCTGCTGTTCAACGGCTGT 125
  QY 805 TACAGCTCTGACATTTCTAGCAGCTATAGATGTAGCGATTCAAGATAAAGTCGATGTTCTT 864
  Db 126 TACAGCTCAGACATTTCTGGCGGATAGACGTAGCGATACAAGACAAAGTCGACGTTCTA 185
  QY 865 TCGCTTTCCCTTGGCGGTTTCCCTATTCTTGTATGATGACACAAATCGCATTTGGACA 924
  Db 186 TCCCTCTCCCTAGCGGTTTCCCTATCCCTTGTATGATGACACTATCGCATTTGGTACC 245
  QY 925 TTCCGAGCATGGAACGCGGTATATCTGTAATCTGTGAGTGTGTAACAACGCTCCCAATC 984
  Db 246 TTCCGCGCCACGGAACATGGAACTCTCGTTGTCGCGGGTGGCAACACGCGCCGCTC 305
  QY 985 GAAAGCTCTGTGTGCAAAACACAGCTCTTGGGTCTTCAACCAATTTGCGCGAGGACGCTTGT 1044
  Db 306 GCTAGCTCGGTGGGAAACACAGCTCTTGGGTCTTCAACCAATTTGCGAGCAGGAACTCGAT 365
  QY 1045 CGAAGATTTCCGCTGTGGTCAGATTAGCCAAACGGAAGCTTCTCTATGAGAGATCATTTG 1104
  Db 366 AGAAAATTTCCCGGGGTTTGTAGGTAGTCCAAACGGAAGCTTACTCTACGAGAGTCTGTTG 425
  QY 1105 TATCCGGAAAAGGTATAAAGATGCCGGGAGAGAGGTTGAGGTGATTTCCTGCACAGGA 1164

```

```

Db      426  TACCAGGAAAGGATTAAGAGAGGCTGAGAGAGATTGGAGATTCTACGTGGGGGA 485
Qy      1165  GAGATTAAGAGAGTGTGTTCTTTGAGAGGTTCTCTCCAAAGAGAAATCCGAGGC 1224
Db      486  GGAGACAAAGAGAGAGATTGTTCTTGAGAGGCTCGTTCCGAGAGAGAGTCCAAAGC 545
Qy      1225  ABAATGGTGATTGTGATCCGGAGTCAATGGAGATCGGAGAGAGAGAGCGGTAAA 1284
Db      546  AAGATGGTGATATCGGACCTGGAGTCAACGGAGATCAGAGAGAGAGAGCGGTAAA 605
Qy      1285  GAAGCTGGAGAGTGAATGATCTTAGCCAAATACAGAGATCAACCAAGAGAGATTCT 1344
Db      606  GAAGCTGGCGGCTGGAGTGAATCTGGCTAAACCGAGATTAACCAAGAGAGAGATTCT 665
Qy      1345  ATTGACGTTTCATCTTTACCAGTACATTGA 1375
Db      666  GTAGACGTTTCATTGATACCAGTACGTTGA 696

RESULT 3
BH998381/c
LOCUS      oeh58ell.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
DEFINITION
ACCESSION  BH998381
VERSION    BH998381.1 GI:23544689
KEYWORDS   GSS.
SOURCE     Brassica oleracea
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 676)
AUTHORS    Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE      Whole genome shotgun reads from Brassica oleracea
JOURNAL    Unpublished
COMMENT    Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: oeh58 row: e column: 11
Seq primer: -21UPpOT forward
Class: shotgun
High quality sequence start: 14
High quality sequence stop: 551.
FEATURES   source
            location/Qualifiers
            1..676
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /db_xref="taxon:3712"
            /clone_lib="B.oleracea002"
            /notes="vector: pOTw13; Whole genome shotgun library from
            flowering buds. DNA was purified from a crude nuclear
            prep using Brassica oleracea TO1000DH3 buds provided by
            Thomas Osborn at the University of Wisconsin. Genomic
            DNA was provided by Pablo Rabinowicz (CSHL) and the
            shotgun library prepared at Washington University Genome
            Sequencing Center."
BASE COUNT 138 a 177 c 208 g 153 t
ORIGIN

Query Match      20.3%; Score 472.4; DB 28; Length 676;
Best Local Similarity 81.3%; Pred. No. 5,7e-104;
Matches 548; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Qy      351  CAAGTCTCTGGAGCTCAGCGGTTTTGAAACTCCGGTGATGCTCTAAATCTCGGTTGG 410
Db      676  CAAGTCTTAGGACTCAACGGTCAGGACCTCAGCGTATGCTCTAAATCCCGGTCGG 617
Qy      411  TCAGGCACAATATCCGCGCTGTTGATCTGGAGTTGGCCCTGAAAGTCTAGCTTTGA 470

```

```

Db      616  CCAAGGCACGATCATCGCGTTCCTCGACACAGGAGTCTGGCCCGAAAGCCCTAGCTTGA 557
Qy      471  CGATACCGGAATGCTTCGATTCCACGGAATGGAAGGATTTGCCAAGAGAGAGAAAG 530
Db      556  CGACACAGGGGATGCTTCAGTCCCAAGCAAATGGAAGGTGTTTTGCCAAGAGAGAGAG 497
Qy      531  TTTCAAGTCTTCGAGCTGTAAACCGGAAGCTAAATCGGTGCTAGATTCTTCTATCAGAGACA 590
Db      496  CTTTCACTTCTCAAACTGTAAACAGAAACTAATCGGCGCTAGATTCTTCTATCAGAGAGCA 437
Qy      591  CCGTGTGCTGAATTCACAGAGAGAAATCACCAGAAATGCTCGTGAATACATTTCCGCAAG 650
Db      436  CCGGTAGCAACTCTCTTTGGACTCACCAGAAATGCTCGGCAATACATATCGGCACG 377
Qy      651  AGATTCAACGGGACACGGGACTCAACGGCTCAACAGTTGGTGGATCCTCTGTTTCGAT 710
Db      376  AGACTCCACGGGGCAGCGGACCCACACGGGCTCAACTGCTGGCGGCTCCTCTGTTTCGAC 317
Qy      711  GGCGAATGTTCTTGGCAATGGAGCTGGTGGCTCGTGGGATGCTCCTCGAGCTCAGAT 770
Db      316  GGCGAGCGTTCCTCGGTAAACGGAGCGGGGTAGCCCGGGATGGCGCAGAGCTCAGCT 257
Qy      771  TGCAGTCTATAAAGTCTGTTGGTTCAATGGTTGTTACAGCTCTGACATTTCTAGCAGCTAT 830
Db      256  GCGGTTTACAAAGTCTGCTGTTCAACGGCTGTTACAGCTCAGACATTTCTGGCGGCGAT 197
Qy      831  AGATGTAGCGATTCAAGATAAAGTCGATGTTCTTTCGCTTTCCCTTGGCGGTTTCCCTAT 890
Db      196  AGACGTAGCGGATACAAGACAAAGTCGACGTTCTATCCCTCTCCCTAGCGGTTTCCCTAT 137
Qy      891  TCCTTTGTATGATACACATCGGCATGGAACATTCGAGCCATGGAACGCGGTATATC 950
Db      136  CCCTTTGTATGATGACACTATCGCCATTGGTACCTTCCGCGCCACGGAATGGAATCTC 77
Qy      951  TGTATCTGTCCAGCTGCTACACGCTGTAACACGCTGCCAATCGAAAGCTCTGTTCAACACAGCTCC 1010
Db      76  CGTTGCTGCGCGGGTGGCAACACGCGCCCTCGTAGCTCGGTGGCGAACAACAGCTCC 17
Qy      1011  TTGGGTCTCAACCA 1024
Db      16  TTGGGATTCACCA 3

RESULT 4
BH464077/c
LOCUS      BH464077.1 GI:17659406
DEFINITION BOGEF01TF BOGE Brassica oleracea genomic clone BOGEF01, genomic survey sequence.
ACCESSION  BH464077
VERSION    BH464077.1
KEYWORDS   GSS.
SOURCE     Brassica oleracea
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 670)
AUTHORS    Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE      Whole genome shotgun sequencing of Brassica oleracea
JOURNAL    Unpublished
COMMENT    Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES   Location/Qualifiers
            1..670
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
source

```

```
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGEF01"
/clone_lib="BOGE"
/notes="Vector: pHS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHS1 using BstXI linkers."
BASE COUNT 137 a 188 c 186 g 159 t
ORIGIN
Query Match 19.8%; Score 461; DB 28; Length 670;
Best Local Similarity 80.6%; Pred. No. 3.4e-101; Indels 0; Gaps 0;
Matches 539; Conservative 0; Mismatches 130;
QY 472 GATACCGGATGCTTCGATTCCACGGAATGGAAGGATTTGCCAAGAGGAGAAAGT 531
DB 670 GACACAGGATGCTTCAGTCCCAAGCAATGGAAGGTTTGCCAAGAGGAGAGC 611
QY 532 TTCACTTTCGAGCTGTAACCGGAGCTAATCGGTGCTAGATTCTTCATCAGAGACAC 591
DB 610 TTCACTTTCGAGCTGTAACCGGAGCTAATCGGTGCTAGATTCTTCATCAGAGGCCAC 551
QY 592 CGTGTGCTAATTCACAGAGGATCACCAGATCGCTCGTCAATACATTTCCGCAAGA 651
DB 550 CGGTAGCAACTCTCTTTGGACTCACCAGATCGCTCGGCAATACATATCGGACGA 491
QY 652 GATTCAACGGGACACGGGACTCACACGGCTCAACAGTTGGTGGATCTCTGTTTCGATG 711
DB 490 GATCCACGGGACACGGGACCCACACGGCTCACTGCTGGGGCTCTCTGTTTCGAGC 431
QY 712 GCGAATGTTCTTGGCAATGAGCTGTGTGGTCTGTGGAGTCTCTCGAGCTCACATT 771
DB 430 GCGAGCGTTCTCGTTAACGAGCGGGTAGCCCGGGATGGCGCAGGAGCTCACGTG 371
QY 772 GCGAGCTATTAAGTCTGTTGGTCAATGTTGTTTACAGCTCTGACATTTACAGCTATA 831
DB 370 GCGGTTTACAAAGTCTGCTGTTTCAACGGCTGTACAGCTCAGACATTTGCGCGGATA 311
QY 832 GATGTAGCGATTCAAGATAAGTCGATGTTCTTTCCTTTCCTTGGCGGTTTCCCTATT 891
DB 310 GAGTAGCGGATACAGACAAAGTCGAGCTTCTATCCCTCTCCAGGCGGTTTCCCTATC 251
QY 892 CTTTGTATGATGACACAATTCGCAATTCGAGCCATTCGAGCCATGGAACGCGTATATCT 951
DB 250 CTTTGTATGATGACACTATCGCAATTCGCAATTCGAGCCATTCGAGCCATTCGCAATCTCC 191
QY 952 GTAATCTGTGACGCTGTAACACGGTCCATCGAAAGCTCTGTTCGCAACACAGTCTCT 1011
DB 190 GTTGTCTGTGCGGGTGCGCAACACGGCGGCTCGTGTGCTGTGCGCAACACAGTCTCT 131
QY 1012 TGGGCTCAACCATTCGCGCAGCAGCTGTGATCGAAGATTTCCCGCTGTGCTCAGATTA 1071
DB 130 TGGGCTCAACCATTCGCGCAGCAGCAGCTGTGATCGAAGATTTCCCGGGGTTGTTAGGTTA 71
QY 1072 GCCAACGGAAAGTTCTCTATGGAGAGTCAATGTATCCGGGAAAAGGTATAAGAAATGCC 1131
DB 70 GCCAACGGAGAGTACTCTACGAGAGTCTGTGTACCCAGGAAAGGATAGAGGCGA 11
QY 1132 GGGAGAGAG 1140
DB 10 GAGAGAGTG 2

RESULT 5
BH599290/c 635 bp DNA linear GSS 15-DEC-2001
LOCUS BH599290 BOWR58TR BOWR Brassica oleracea genomic clone BOWR58, genomic
DEFINITION survey sequence.
ACCESSION BH599290
VERSION BH599290.1 GI:17851736
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eucosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 635)
TOWN C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BOWR58TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
source
1..635
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGEF01"
/clone_lib="BOGE"
/notes="Vector: pHS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHS1 using BstXI linkers"
BASE COUNT 127 a 198 c 162 g 148 t
ORIGIN
Query Match 18.8%; Score 438.2; DB 28; Length 635;
Best Local Similarity 80.6%; Pred. No. 1.2e-95;
Matches 512; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
QY 635 AATACATTTCCGCAAGATTCACCGGACACCGGACTCACCGCTCACCGCTCACAGTTGGTG 694
DB 635 AATACATTCGCGACGAGACTCCACGGGCCACCGGACCCACACCGCTCACTGCTGCGG 576
QY 695 GATCCTCTGTTTCGATCGGCAATGTTCTTGGCAATGAGCTGTGTGGCTCGTGGGATGG 754
DB 575 GCTCCTCTGTTTCGACGGCGAGCGTTCTCGGTACGAGCGGGGTAGCCCGGGATGG 516
QY 755 CTCCTGGAGCTCAATTCGAGTCTATAAGTCTGTGGTTCATGTTGTATTACAGTCTG 814
DB 515 CGCCAGAGCTCAGCTGCGGTTTACAAAGTCTGTGTTCAACGGCTGTATTACAGTCTAG 456
QY 815 ACATTCAGACGCTATAGATGTAGCGATTCAAGATAAGTCTGATGTTCTTTCGTTTCCC 874
DB 455 ACATTCGCGCGGATAGACGTAGCGATACAGACAAAGTCGACGTTCTATCTCCTCTCCC 396
QY 875 TTGCGGCTTTCCTATTCCTTTGATGATGACCAATTCGCCATTCGCAATTCGAGCCCA 934
DB 395 TAGCGGTTTCCCTATCCCTTTGATGATGACACTATCGCCATTCGATGTAACCTTCGCGCCA 336
QY 935 TGGAAACCGGTATATCTGTAATCTGTGAGCTGGTAAACAAACGGTCCCAATCGAAAGTCTG 994
DB 335 CGGAACATGAAATCTCGTTGTTCTGCGCGGTGCAACACGCGCCGCTCGTAGTCTCG 276
QY 995 TTGCAAAACAGCTCTCTTGGTCTCAACCATTCGCGCAGCAGCTGTGATCGAAGATTC 1054
DB 275 TGGCAACACAGCTCTCTTGGTCTCAACCATTCGCGCAGCAGCTGTGATCGAAGATTC 216
QY 1055 CCGCTGTGCTCAGATTAGCCCAACGGAAAGCTTCTCTATGGAGAGTCAATGTTATCCGGAA 1114
DB 215 CGGGGTTGTTAGGTTAGCCCAACGGGAAAGTCTCTACGAGAGTCTGTTGATCCCGGGA 156
QY 1115 AAGGTATAAAGATTCGCGGAGAGAGGTTGAGGTGATTTACGTCAACAGGAGAGATAAG 1174
DB 155 AAGGATATAAAGAGCTGAGAGAGGTTGAGGATTTGTTACGTGCGGGAGAGACAAAG 96
QY 1175 GAAGTGAGTTCTGTTTCGAGAGGTTCACTTCCAGAGAGAGAAATCCGAGGCAAAATGTTGA 1234
DB 95 GAACCGAGTTTGTCTTCGAGAGGCTCGCTTCCGAGAGAGAGTCTCCAAAGGCAAGATGTTGA 36
```

QY	1235	TTTGTGATCGCGAGTCATTAATGAAGATCGGAGAA	1269
Db	35	TATGCCACCGTGAGTCAACGGGAGATCAGAGAAA	1
RESULT 6	BH597572	627 bp DNA linear GSS 15-DEC-2001	
LOCUS	BOHON29TF BOHO Brassica oleracea genomic clone BOHON29, genomic		
DEFINITION	survey sequence.		
ACCESSION	BH597572		
VERSION	BH597572.1 GI:17850024		
KEYWORDS	GSS.		
SOURCE	Brassica oleracea		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.		
REFERENCE	1 (bases 1 to 627)		
AUTHORS	Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.		
TITLE	Whole genome shotgun sequencing of Brassica oleracea		
JOURNAL	Unpublished		
COMMENT	Other GSSs: BOHON29TR Contact: Chris Town		
TIGR	9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdtown@tigr.org		
Seq primer:	TF		
Class:	sheared ends.		
FEATURES	Location/Qualifiers		
source	1..627		
/organism=	"Brassica oleracea"		
/mol_type=	"genomic DNA"		
/strain=	"TO1000DH3"		
/db_xref=	"taxon:3712"		
/clone_lib=	"BOHON23"		
/note=	"Vector: pHSO1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHSO1 using BstXI linkers"		
BASE COUNT	161 a 186 c 146 g 134 t		
ORIGIN			
Query Match	18.3%; Score 426.8; DB 28; Length 627;		
Best Local Similarity	81.2%; Pred.No. 78-93;		
Matches	508; Conservative. 0; Mismatches 117; Indels 1; Gaps 1;		
QY	148	CITTCCTTTCTCAAGAAGCGGTTTTAGTGTTGAAGAAGAAGAGAGACCTTCTTCT	207
Db	1	CITTCCTTTCTCAAGAAGCTGTTCTAGTATTGAAGAAGAAGAGACCTTCTACA	60
QY	208	CGACTTCTTACTCCATGGCTTCGATGAAGATTGCTGCATGACTGAATCA	267
Db	61	AGGATTCTTACTCTTACGCCCTTCGCTTGAAGGATTCAGCTCAGTAAACCGAATCA	120
QY	268	GAAGCCGAGATCTAGATAATTCACCTGAGTTGTTGAGTGAAGACCTGACCATTCTT	327
Db	121	GAAGCCAAACACTCAAGAAGCTTACCTGAAGTTGTTGAGTGAAGACCTGACCATTCTC	180
QY	328	CAGGTTCAACACCTTACTCTTACAGTTCTTGGGACTCGACGGTTTGGAAATCCCGGT	387
Db	181	CAAGTCAAACCACTTACTCTTACAGTTCTTAGGACTCAACGGTCCAGGACCTCCAGC	240
QY	388	GTATGTCTAAATCTCGGTTTGGTCAAGGCACAAATATCGGCGTCTTGTATCATCGAGTT	447
Db	241	GTATGTCTAAATCCCGGTCGCGCCAAGGCAAGATCATCGGGTTCTCGACACAGGAGTC	300
QY	448	TGCGCTGAAAGTCTTACGATTTGACGATACCGGAATGCCTTCGATTCACGGAATTGGAAA	507
Db	301	TGCGCCGAAAGCCCTTACTCTGACGACACAGGATGCTCTTACGTCCTCCCAAGCAATTGGAAA	360

QY	508	GGGATTGCCAAGAAGAGAAAGTTTTCAGTTCTTCGAGCTGTAAACCGGAGCTAATCGGT	561
Db	361	GGTGTGTTGCCAAGAAGAGAGAGCTTTCATCTTCCTCACTCTAAACAAGAAAACTAATCGGC	422
QY	568	GCTAGATTCTTCATCAGAGGACA-CCGNTGCGTAATTCCACAGAGGAATCACCAAACAT	626
Db	421	GCTAGATTCTTCATCAGAGGCCCCCGGTAGCCAACTCTCTTTGGACTCACCAAACAT	480
QY	627	GCTCTGCTGAATACATTTCCGCAAGAGATTCAACGGGACACGGGACTCACACCGCTCAAC	686
Db	481	GCTCTGCGAATACATATCGCACAGAGATCCACGGGGCAGCGGACCCACACCGCTCAAC	540
QY	687	AGTTGGTGGAATCCTCTGTTTCGATGCGGAATGTTCTTTGGCAATGAGAGTGTGTGGCTCG	746
Db	541	TGCTGCGGCTCTCTCTGTTTCGACGGGAGGCTTCTCGTAAACGAGAGCCGCTAGCCCG	600
QY	747	TGGATGAGCTCTCTGAGGCTCACAATG 772	
Db	601	CGGATGCGGCCAGGAGCTCAGTGG 626	
RESULT 7	BH442444/c	640 bp DNA linear GSS 12-DEC-2001	
LOCUS	BOHGT85TR BOHG Brassica oleracea genomic clone BOHGT85, genomic		
DEFINITION	survey sequence.		
ACCESSION	BH442444		
VERSION	BH442444.1 GI:17628158		
KEYWORDS	GSS.		
SOURCE	Brassica oleracea		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.		
REFERENCE	1 (bases 1 to 640)		
AUTHORS	Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.		
TITLE	Whole genome shotgun sequencing of Brassica oleracea		
JOURNAL	Unpublished		
COMMENT	Other GSSs: BOHGT85TF Contact: Chris Town		
TIGR	9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdtown@tigr.org		
Seq primer:	TR		
Class:	sheared ends.		
FEATURES	Location/Qualifiers		
source	1..640		
/organism=	"Brassica oleracea"		
/mol_type=	"genomic DNA"		
/strain=	"TO1000DH3"		
/db_xref=	"taxon:3712"		
/clone_lib=	"BOHGT85"		
/note=	"Vector: pHSO1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHSO1 using BstXI linkers"		
BASE COUNT	149 a 211 c 126 g 154 t		
ORIGIN			
Query Match	18.2%; Score 424; DB 28; Length 640;		
Best Local Similarity	78.9%; Pred.No. 3.4e-92;		
Matches	505; Conservative 0; Mismatches 135; Indels 0; Gaps 0;		
QY	1172	AAGGAAGTGAAGTTCTGTTTGAAGGCTCACTTCCAAAGAGAGAAATCCGAGCAAAATGG	1231
Db	640	AAAGAAAGCGAGTTGTTCTTGAGAGGCTCGTCCCAGAGAGAGTGTCCAAAGCAAGATGG	581
QY	1232	TGATTTGTCATCGCGAGTCAATGGAAGATCGGAGAAAGAGAGCGGTTAAGAGCTG	1291
Db	580	TGATATGCCACCGCTGGAGTCAACGGGAGATCAGAAAGAGGACAGCGGTTAAGAGCTG	521

QY 1292 GAGGAGTTGCAATGATCTTACCAATACAGAGATCAACCAAGAGAGATCTTATGACG 1351
 DB 520 GCGGCGTGGGATGATCTTGGCTACCAACGAGATTAAACAGAGAGAGATCTTATGACG 461
 QY 1352 TTCACTCTTACAGCTACATGATGTTGTTACACTAGTCACTCTTCTGAAGGCTTATG 1411
 DB 460 TTCACTTATGATACCAAGTACGTTGATGTTGTTACAGAGATCTGTTGTTTGAAGGTTACG 401
 QY 1412 TTAATCCCGGTTGAACCAAGCGCGGATATTTTGGTGTAGCTGATTTGGGAGT 1471
 DB 400 TGAGGACACCGGTAAGACCGAAAGCTAGGTTAATTTTCGGCGGACCGTGTATGGAGGT 341
 QY 1472 CACGAGCACCGGAGGTGGCTCAGCTTTTCAGCTCGAGGACCGAGTTTATGCAATCTTCGA 1531
 DB 340 CGAGAGGGCTTGAAGTGGCGCAGTTTCTCAGGAGAGGACCGAGTTTGGCTAACCTTCGG 281
 QY 1532 TACTAAACCGGATATGATGCTCCGGGAGTCAATATCATATGCGGCTTGGCTCAAAATC 1591
 DB 280 TGCTTAAACCGGATTTGATCGCGCAGGTGTCAATATTATAGTGTGCTGGCTTCAGAAATC 221
 QY 1592 TAGACCAACCGGACTCTTATGATCAAGAGAGTTAACTTCACTGTAATGTCAAGAA 1651
 DB 220 TTGACCGACCGGGCTTCCCTTATGATTCGAGGAGGTTAACTTCACTGTAATGTCAAGAA 161
 QY 1652 CTTCAATGCTTGTCCACATGTTAGCGGATCACTGCTTATCCGGTCTGCATACCCGA 1711
 DB 160 CTTGATGCTTGTCCACATGTTAGCGGATCACTGCTTATCCGGTCTGCATACCCGA 101
 QY 1712 ACTGGTCTCAGCTGCAATCAATCCGCAATGATGATCAAGAGGATTTGTACGATCGTC 1771
 DB 100 GCTGGTCTCAGCTGCTATCAGATCAGGAGTATGATGACGCGGTGATTTGTATGATCGGA 41
 QY 1772 AAGGAAAGCGATAAGGATGTTAACAACCAACCGCGGTGT 1811
 DB 40 GAGGAAAGAGATTAGGACCGGGACAAACCGCGGGAGT 1

RESULT 8
 BH601308/c
 LOCUS BOHS89TF BOHS Brassica oleracea genomic clone BOHS89, genomic
 survey sequence.
 ACCSSION BH601308
 VERSION BH601308.1 GI:17853754
 KEYWORDS GSS.
 SOURCE Brassica oleracea

ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished
 Other_GSSs: BOHS89TR
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES
 source
 1..803
 Location/Qualifiers
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="T0100DH3"
 /db_xref="taxon:3712"
 /clone="BOHS89"
 /clone_lib="BOHS"
 /note="Vector: phos1; Site_1: BstXI; 2-3 kb sheared

BASE COUNT 245 a 223 c 121 g 214 t
 ORIGIN
 Query Match 17.6%; Score 409.2; DB 28; Length 803;
 Best Local Similarity 80.6%; Pred. No. 1.4e-88;
 Matches 493; Conservative 0; Mismatches 113; Indels 6; Gaps 1;
 QY 1723 GCTCAATCAAAATCCGATTCAGCAACAGCGGATTTTACGATCGTCAAGGAAAGCG 1782
 DB 803 GCTGCTATCAGATCAGGATGATGACGCGCTGATTTGATGATCGGAGGAAAGAG 744
 QY 1783 ATAAAGGATGGTAACAAACAGCGCGGTGTTTTCGATTTGGAGAGGCGATGTAAATCCG 1842
 DB 743 ATTAGGACCGGGACAAACCGCGGGAGTGTGTTGCTATTGGAGCAGGCAATGTAAATCCG 684
 QY 1843 CAAAGGCGATAACCCGGGATTTGTTTACAACATTCACAGTGGATACATAACTTAC 1902
 DB 683 GTTAAGCGGATTAACCCGGGGTGGTTTACAACATTCACCGGTTGATTACATTCGTTAC 624
 QY 1903 CTCTGCACTCTTGGATTCACAGATCAGATATTTTACCAATCACTCATAGAAACGTGAGC 1962
 DB 623 TTATGCACTATGGGTTTACTAGATCGATATTTTAGCGATCACTCACAGGACGTAGC 564
 QY 1963 TGCAATGAAATATTGCGGAAACCCGGGTTTATGCTCAATTAACCGTCTGATAGCGGTG 2022
 DB 563 TCGGTGCGGTACTGAGGAGAGCCCGGTTTATGCTTAAATATCGTCTATTTCGGTT 504
 QY 2023 ATTTTCAAAAGTGCAAGACTACGAGATGATCACAAGGCGTGTCACTACGTTGGAGT 2082
 DB 503 ATTTTACGCTGGGAGGACTAAGGAGATGATCAGAGGCGGTGTGACTAAACGTTGGAGT 444
 QY 2083 CCTAATCGATATCACTCAGTGAATGTCAAGGCTCCAGAGGGGATCAAAGTTATTGTCAAT 2142
 DB 443 CCTAATCGGTGTACACTGTGAATGTGAAGCTCCTATGGGATTAATGTGATTTGAAG 384
 QY 2143 CCTAAGAGACTTGTGTTCAAACACGCTGATCAGAGCTGAGCTATAGAGTATGATGTTGTA 2202
 DB 383 CCTAAGAGGCTTGTGTTTATGCTACGTTGATCAACAGTTGAGCTATAGAGTTGTTGTTG 324
 QY 2203 TTGAAGAGAAACAAACAGAGAGG-----GAAGTGGCTAGCTTTGACAGAGGCGAGTTG 2256
 DB 323 TTGAAGAGGAGGAGAGCAGCAGAGAGAGAGAGTGGTGGTACCTTTGCGGATGACAGATTG 264
 QY 2257 ACTTGGGTCACTCTCATAATCTGATCAGCGAGTTAGAGTCCAACTCTGTAAACCTTG 2316
 DB 263 ACTTGGGTCACTCTCGTGTTCGATCAGCGAGTTAGGAGTCCAACTCTCTGTACCTTG 204
 QY 2317 AAGACTAACTGA 2328
 DB 203 AAGAACTATTGA 192
 RESULT 9
 BH714875/c
 LOCUS BOMLB06TF BO_2_3_KB Brassica oleracea genomic clone BOMLB06,
 genomic survey sequence.
 ACCESSION BH714875
 VERSION BH714875.1 GI:18809000
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished
 Other_GSSs: BOMLB06TR
 Contact: Chris Town
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TP
 Class: sheared ends.

FEATURES

Location/Qualifiers
 1. .527
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOML806"
 /clone_lib="BO.2.3 KB"
 /notes="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHOS1 using BstXI linkers"
 155 a 157 c 88 g 127 t

BASE COUNT

155 a 157 c 88 g 127 t

ORIGIN

Query Match 15.3%; Score 356.2; DB 28; Length 527;
 Best Local Similarity 81.2%; Pred. No. 1e-75;
 Matches 428; Conservative 0; Mismatches 93; Indels 6; Gaps 1;
 QY 1804 GCCGGTGTTCGGATTGGAGCAGGCGATGTAATCCGAAAGCGATAAACCCGGA 1863
 Db 527 GCGGGAGTGTTCGATTGGAGCAGGCGATGTAATCCGTTAAGCGATTAAACCCGGG 468
 QY 1864 TTGGTTTACACATTCACACAGGATTACATAACTCTCTGCACTCTGGATTACAA 1923
 Db 467 TTGGTTTACACATTCACACCGGTGATTACATTCCTTATGCACTATTGGTTTACT 408
 QY 1924 AGATCAGATATTTAGCAATCACTATAGAACGTCAGTGTCAATGCAATATTGGCGAAA 1983
 Db 407 AGATCGATATTTAGCGATCACTACAGAAAGTGTAGTGGTGGTCTAGGAGAG 348
 QY 1984 AACCCGGTTTTAGTCTCAATTAACCGTCGATAGCGGTGATTTTCAAGGTGGCAAGCT 2043
 Db 347 AGCCCGGTTTTAGTCTTAATATTCGCTATTTCGGTTATTTTATAGAGTGGAGGACT 288
 QY 2044 ACGGAGATGATCACAGCGGTGCTACTAAGTTGGGAGTCTTAACCTCGATATACTCAGTG 2103
 Db 287 AAGGAGATGATCACAGCGGTGCTACTAAGTTGGGAGTCTTAACCTCGGTGTACACTGTG 228
 QY 2104 AATGCTAAGCTCCAGAGGGGATCAAGATTATTTGTCATCTCAAGAGACTTGTGTTCAAA 2163
 Db 227 AATGTTAAGCTCTTATGGGATTAATGTAATTTAAGCCTAAGAGGCTTGGTTTACT 168
 QY 2164 CACGTGATCAGACGCTAGCTATAGAGTATGTTTGTATTGAAGAGAAACACAGAGGA 2223
 Db 167 CACGTTGATCAACGTTGAGCTATAGAGTTGGTTTGTGTTGAAGAGGGAAGCAGACA 108
 QY 2224 GG-----GAAGGTGGTACGTTTGGCAAGGGCAGTTGACTTGGGTCAACTCTCATAAT 2277
 Db 107 GGAGAGAAGGTGGTGTGCTTCTTGGGATGGACAGTTGACTTGGGTCAACTCTCGTGAT 48
 QY 2278 CTGATGACGCGAGTTAGATCCCAATCTCTGTAACCTTGAAGACTAA 2324
 Db 47 TCGATGACGCGAGTTAGAGTCCCAATCTCTGTGACCTTGAAGAATCA 1

RESULT 10

BH793726
 LOCUS BH793726 767 bp DNA linear GSS 02-APR-2002
 DEFINITION ME MBa0001K22f Manihot esculenta Manihot esculenta genomic clone
 accession BH793726
 accession BH793726.1 GI:19891014
 keywords GSS.
 source Manihot esculenta (cassava)
 organism Manihot esculenta
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

REFERENCE 1 (bases 1 to 767)
 AUTHORS Tonkins,J.P., Peregine,M., Main,D., Goicoechea,J.L., Blackmon,B.,
 Atkins,M., Tohme,J. and Wing,R.A.
 TITLE New Genomic Resources for Cassava (Manihot esculenta): Development
 of a Deep-Coverage BAC Library and Preliminary STC Analysis
 JOURNAL Unpublished
 COMMENT Contact: Tonkins J
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
 Tel: 864 656 6419
 Fax: 864 656 4293
 Email: jtmkns@clemson.edu
 Total High Quality bases = 313
 Seq primer: TAATACGACTCACTATAGGG
 Class: BAC ends
 High quality sequence start: 77
 High quality sequence stop: 546.
 Location/Qualifiers

1. .767
 /organism="Manihot esculenta"
 /mol_type="genomic DNA"
 /strain="MECW72"
 /db_xref="taxon:3983"
 /clone="ME_MBa0001K22f"
 /tissue_type="Leaf"
 /lab_host="E. coli"
 /clone_lib="Manihot esculenta"
 /note="Vector: pCUGIBAC-1; Site 1: HindIII; Site 2: NotI;
 For more details on library preparation and sequence
 analysis see
 http://www.genome.clemson.edu/projects/etc/cassava/ME_MBa
 To order clones from this library see
 http://www.genome.clemson.edu/orders "
 233 a 183 c 153 g 186 t 12 others

FEATURES

source
 1. .767
 /organism="Manihot esculenta"
 /mol_type="genomic DNA"
 /strain="MECW72"
 /db_xref="taxon:3983"
 /clone="ME_MBa0001K22f"
 /tissue_type="Leaf"
 /lab_host="E. coli"
 /clone_lib="Manihot esculenta"
 /note="Vector: pCUGIBAC-1; Site 1: HindIII; Site 2: NotI;
 For more details on library preparation and sequence
 analysis see
 http://www.genome.clemson.edu/projects/etc/cassava/ME_MBa
 To order clones from this library see
 http://www.genome.clemson.edu/orders "
 233 a 183 c 153 g 186 t 12 others

BASE COUNT

233 a 183 c 153 g 186 t 12 others

Query Match 14.0%; Score 326.8; DB 28; Length 767;
 Best Local Similarity 68.4%; Pred. No. 1.6e-66;
 Matches 473; Conservative 0; Mismatches 217; Indels 2; Gaps 2;

QY 1400 TGAAGGCTTATGTTAATGCGACCGGTGAAACCAAGGCGCGGATAATTTTGGTGGTACGG 1459
 Db 70 TAAATCTTACATTAATCTACAAAGTAAACCAAGAGCTCGAATCATTTTCAAGGGAACGTG 129
 QY 1460 TGATGGGAGTCACGAGCAGCGGAGTGTCTAGTTTTCAGCTCGAGGACCGAGTTTAG 1519
 Db 130 TAAATCGGAAATCCAGAGCACCAGAGTAGCTCAGTTTTCAGTAGAGGACCAAGTTTAA 189
 QY 1520 CCAATCCTTCGATACTAAACCGGATATGATTGCTCCGGAGTCAATATCATTCGCGGCTT 1579
 Db 190 CTACCCCTTCAATCTCTAAACCGATGTAATTGCACTGGAGTAAACATCATTCGAGCTT 249
 QY 1580 GGCCTCAAAATCTAGGACCAACCGGACTTCCTTATGATTTCAAGAGAGTTAACTTCACTG 1639
 Db 250 GGCCTCAAAACTTAGGCCCACTGGTCTTCCAGAAGATCCAAAGAGAGTGAACCTTCACAG 309
 QY 1640 TAATGTAGGAACCTTCAATGTCTTGTCCACATGTTAGGGAATCACTGCTTCTTATCCGGT 1699
 Db 310 TAATGTAGGAACATCCATGCTGCTTGTCTCATGTGAGTGAATGCTGCCCTGATCCGCT 369
 QY 1700 CTGCATACCCGAACCTGGTCTCCAGTGCATCAATCAATCGCATTTGATGACACAGCGGATT 1759
 Db 370 CAGCTCATCCCGGATGGACACTCTGCGAGTTAATATGAGCAATATGACTACTGCAGATA 429
 QY 1760 TGTACGATCGTCAAGGGAAGCGATAAGGATGTTAAACAAACCCAGCCGGTGTGTTGGGA 1819
 Db 430 TAACTGACCATTCCTGGAATCCCAATATTGGATGTTGACAAACACGACGACTCTTTGCAA 489
 QY 1820 TTGGAGCAGG-GCATGTGAATCCGCAAAAGCGGATAAACCCCGGATTTGGTTTACACATT 1878
 Db 490 TTGGAGCCTGGCATATACACCCAGAGAGGCCATCAATCCGAGATTGATATGATATCATC 549

```

QY 1879 CAACCAAGTGAATACATAACTTACCTCTGCACTCTTTGATTCACAAAGATCAGATATTTTA 1938
DB 550 AGCCCCAGAGAGATGTCACCCACCTATGNCACCTAGATACANAGATCAGAAATTTTC 609
QY 1939 GCAATCACTCATAGAACGTGAGCTCAATGGAATATTCGGAAACCCGGGTTTGTAGT 1998
DB 610 ACAATCACACACAGGAATGTGAGCTGNCATGAACTATGCAATGAATAAGGGTTNCAGC 669
QY 1999 CTCAAATTCACCGTCGATGCCGTGATTTTCAACGCTGGCAAGACTACCGAGATGATCACA 2058
DB 670 CTCAAATNATCCTTNCACCTTCTGATNTAAAGCG-GGGATGACNNAGTATGATCCGA 728
QY 2059 AGCGGTGTCACTAACGTTGGAGTCTCAATC 2090
DB 729 AGAGCACTTACCAATGTGGTATCCTAAATC 760

RESULT 11
BF052773 568 bp mRNA linear EST 07-MAR-2003
LOCUS EST438003 potato leaves and petioles Solanum tuberosum cDNA clone
DEFINITION CSTB32P4 5' sequence, mRNA sequence.
ACCESSION BF052773
VERSION BF052773.1 GI:10806669
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
1 (bases 1 to 568)
van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,
Utterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Romming
,C.M., Fry,W.E., Tanksley,S.D. and Baker,B.
TITLE Generation of ESTs from potato leaves and petioles
JOURNAL Unpublished
COMMENT Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/.

FEATURES
Location/Qualifiers
source
1..568
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="CSTB32P4"
/tissue_type="leaflets and petioles"
/dev_stage="8 weeks old plants"
/lab_hosts="SOLR"
/clone_lib="potato leaves and petioles"
/notes="vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Tissue was supplied by Dr. Fry (Cornell University).
Leaflets and petioles were isolated from 8 week old
greenhouse grown plants. The plants were watered and
fertilized freely. The tissue was immediately frozen in
liquid nitrogen."
BASE COUNT 176 a 123 c 133 g 136 t
ORIGIN

Query Match 12.7%; Score 296.8; DB 10; Length 568;
Best Local Similarity 70.4%; Pred. No. 3e-61;
Matches 397; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 1205 CAAGAGAGAAATCCGAGGCAAAATGGTGATTTGTATCCGGAGTCAATGGAGATCGG 1264
DB 4 CAAAGCACAAATCCGAGGCAAAATGGTGATTTGTATCCGGAGTCAATGGAGGCGAG 63
QY 1265 AGAAGAGAGACGGGTAAAGAGAGCTGGAGGAGTTCGAATGATCTTAGCCATACAGAGA 1324

```

```

DB 64 AAAAAGCCAGGTGTGAAGGAGCGAGGTGCTGCCATGATCTTAGCAATACAGCA 123
QY 1325 TCAACCAAGAAGAAGATTCTATTGAGCTTCATCTTTACCAGCTACATTGATTGGTTACA 1384
DB 124 TAAATATGAGGAAGATTCCGTTGATGTCATGTCCTCCAGCAACGTTGATTGGCTCG 183
QY 1385 CTGAGTCAGTCCTTCCTTGAAGGCTTATGTTAATGCCACGGTGAAACCAAAAGGCCGGATAA 1444
DB 184 ATGAATCAATTCAATTACAAAACCTACCTGAACCTCAACAAAAGACCAACAGCTCGATTCA 243
QY 1445 TTTTGTGTGATCGGTGATTTGGGAGGTCACGAGCAGCGGAGGTGGCTCAGTTTTAGCTC 1504
DB 244 TATTGGAGGACCGTAATAGSAAAGCTTAGAGACCTCGAGTAGCTCAGTTTTCTGTCAA 303
QY 1505 GAGGACCGAGTTTAGCCAATCTTCGATACTATAAACCGGATATGATTTGCTCCGGAGTCA 1564
DB 304 GGGGCCCAAGCTATACGATCTTCAATCTCAAACTGATTTGATTTGCTCCAGGGTAA 363
QY 1565 ATATCATTTGGGGTTGGCTTCAAAATCTAGGACCAACCGGAGCTTCCTTATGATCAAGAA 1624
DB 364 ACATAATTGGCGTTGGCCACAAAACCTTAAGCCCGAGTGTCTTCTGAAGATTACGAA 423
QY 1625 GAGTTAACTTCACTGTAATGTCTAGGACCTTCAATGTCTTCTCCACATGTTAGCGGATCA 1684
DB 424 GAGTAAATTTCACTGTTATGTCTAGGGACCTCAATGGCATGTCTCTATGTAAGTGAATTG 483
QY 1685 CTGCTCTTATCCGCTCTGCATACCCGAACCTGCTCCAGCTGCAATCAAAATCCGCATTGA 1744
DB 484 CCGATTGTCTCCATTCAGCTCATCTCTAATGGAGCTCCAGGAGCAATAAGATCCGCATAA 543
QY 1745 TGACAACAGCGGATTTGTACGATC 1768
DB 544 TGACCACTGCAGATACAGCTGATC 567

RESULT 12
BH601319 772 bp DNA linear GSS 15-DEC-2001
LOCUS BOHSA89TR BOHS Brassica oleracea genomic clone BOHSA89, genomic
DEFINITION survey sequence.
ACCESSION BH601319
VERSION BH601319.1 GI:17853765
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
TITLE ; eurosids II; Brassicales; Brassicaceae; Brassica.
JOURNAL 1 (bases 1 to 772)
COMMENT Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BOHSA89TF
Contact: Chris Town
TIGR Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
source
1..772
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TOL000DH3"
/db_xref="taxon:3712"
/clone="BOHSA89"
/notes="vector: pBstXI; 2-3 kb sheared
genomic DNA inserted into pBstXI linkers"

```


BASE COUNT	243 a	207 c	105 g	217 t
ORIGIN				
Query Match	11.7%;	Score 273;	DB 28;	Length 772;
Best Local Similarity	80.9%;	Pred. No. 1.9e-55;		
Matches 318;	Conservative 0;	Mismatches 75;	Indels 0;	Gaps 0;
QY	1	ATGGAACCCAAACCTTTCTTCTCGCATTAATCACTTTCTTCCTATTCTTTGTCTTTCTTCCTGCTCA	60	
Db	380	ATGGAACCTAAGACTCTCTTTTGTCCTCACTATCTTTCTCCTCTTTGTTGTTCTTCTTCTCTCA	439	
QY	61	GAGATCTGCAGAAGCAGACTTATCTGTTTCAGCTTCATCTCTATAGCGAAACGGCTAAA	120	
Db	440	GAACACCCTTAAGAACAGACTTAGCTCAITTCAGCTTCACCTTAATAGCAATTCAGCTAAA	499	
QY	121	ACCTTTTGCTCAAAGTTTGATTGGCATCTTTCTTTCTCCCAAGACGGTTTTAGGTGTT	180	
Db	500	GCTTTTCTTCAAAGCTTGATGGCACTTTCTTTCTTCAAAGAGCTGTTCTAGGTATT	559	
QY	181	GAAGAAGAAGAGGAGACCTTCTTCTCGACTTCTTCTACTCTATGCCATTCGCGATTGAA	240	
Db	560	GAAGAAGAAAACGAAGACCCCTTCTACAAGGATTTCTTACTCTTACGCCCTCTGCCATTGAA	619	
QY	241	GGATTTGCTGCTCAGTTGACTGATCAGAAGCCGAGATCTGAGATATTACCTTGAAGTT	300	
Db	620	GGATTTCTCAGCTCAGTTTAACCGAATCAGAAGCCAAAAACACTCAAGAACTTACCTGAAAGT	679	
QY	301	GTTCCAGTGAGACCTGCACCATGTTCTTCCAGGTTCGAACCACTTACTCTTACAAGTTCTTG	360	
Db	680	GTTGCGATPAGACTGACCATGTTCTCCCAAGTCAACCACTTACTCTTACAAGTTCTTA	739	
QY	361	GGACTCGACGGTTTTGGAAACTCCGGTGATGG	393	
Db	740	GGACTCAACGGTCCAGGACCCCTCCAGCGTATGG	772	
RESULT 13				
CC340521/c				
LOCUS	CGRAO39TV ZM.0.7.1.5 KB Zea mays genomic clone ZMMBma0374H05,	789 bp	DNA	linear GSS 16-MAY-2003
DEFINITION	genomic survey sequence.			
ACCESSION	CC340521			
VERSION	CC340521.1	GI:30809927		
KEYWORDS	GSS.			
SOURCE	Zea mays			
ORGANISM	Zea mays			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.			
AUTHORS	Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.			
TITLE	1 (bases 1 to 789)			
JOURNAL	Consortium for Maize Genomics			
COMMENT	Unpublished Contact: Cathy Whitelaw			
FEATURES	TIGR			
source	9712 Medical Center Drive, Rockville, MD 20850, USA			
	Tel: 301-838-5843			
	Fax: 301-838-0208			
	Email: whitelaw@tigr.org			
	Seq primer: Tg			
	Class: sheared ends.			
	Location/Qualifiers			
	1..789			
	/organism="Zea mays"			
	/mol_type="genomic DNA"			
	/strain="B73"			
	/db_xref="taxon:4577"			
	/clone="ZMMBma0374H05"			
	/clone_lib="ZM.0.7.1.5"			
	/note=vector: pBCSK; Site 1: HincII; 0.7-1.5 kb methylation filtered genomic DNA library"			

```

Ehrhartoideae; Oryzae; Oryza.
1 (bases 1 to 750)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GCTATTAGGTGACACTATAG
Class: BAC ends
High quality sequence stop: 339.
Location/Qualifiers
1. .750
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSUNB008017r"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/notes="Vector: pBACIndigo; Site 1: EcoRI; Site 2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa.
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT 128 a 254 c 256 g 110 t
ORIGIN
Query Match 10.8%; Score 252; DB 28; Length 750;
Best Local Similarity 60.3%; Pred. No. 2.5e-50;
Matches 449; Conservative 0; Mismatches 292; Indels 3; Gaps 2;
QY 1321 GAGATCAACAGAGAGAGATTCTATTGAGCTTCACCTCTTACAGCTACATGATTGCT 1380
DB 2 GAGATAACAGGAGAGAGACTCGTCGATGTCATGTCCTCGCTGCGAGCTCATCGGC 61
QY 1381 TACACTGAGTCAGTCTCTTGAGAGCTTATGTTAATGCCACGCTGAAACCAAGGCGCG 1440
DB 62 TACAGGAGGCGCTCGAGCTCAAGACTAGTCAGTCCAGCGCGCGCGCTGGCGAGG 121
QY 1441 ATAAATTTTGTGTACGGTGATTGGAGTCAAGACACCGAGGTGGCTAGTTTTCA 1500
DB 122 ATAGTGTTCGGCGGACGCGATTGGCGGACGCGCGCGCGCTGGCGCTGTCTCC 181
QY 1501 GCTCGAGACCGAGTTTAGCCATCTTCATATAACCGATATGATTCGTCGGGA 1560
DB 182 GCGCGCGGCGAGCCCTGACGAACCCCTCGGTGCTCAAGCCGACGTGGTCGCCCGGC 241
QY 1561 GTCAATATCATTTGGGCTTGCCCTCAAAATCTAGGACCAACCGACTTCCTTATGATTC 1620
DB 242 GTGAACATCATCGCGGTGGCCCGGCAACTGGGACCTGCGGGCTGGAGGCGGAGCC 301

```

```

QY 1621 AGAAGAGTTAACTCACTGTAATGTGAGGAACCTTCAATGCTTGTCCACATGTTAGCGGA 1680
DB 302 CGCGCGTCCGACTTCACGGTGTCTCTCGGACATCAATGCGTGCCTGCCGACGTGAGCGGC 361
QY 1681 ATACTGCTCTTTATCCGGTCTCATACCCGAACTGCTCTCCAGCTGCAATCAATCGCA 1740
DB 362 ATCGCGCGCTCATCAGGTTCGGCGACCCCGTCTGTGAGCCCGCGATGGTGGCGTCCGG 421
QY 1741 TTGATGACAAACAGCGGATTGTTACGATCGTCAAGGAGAAAGCGGATAAGGATGTAACAA 1800
DB 422 ATCATGACGACCGCGGACGTGACGACCGGAGGCGAGCCGATCATGACGGGACGGC 481
QY 1801 CCA--GCCGGTGTGTTTTCGATTGAGCAGGCGGATGTAATCCGAAAAAGCGGATAAAC 1858
DB 482 GCAAGGCCGACGCGTACGCGCATGCGCGCGGCGCACGTGAANCCGCGAGGCGCGTGCACC 541
QY 1859 CGGATTTGGTTTACACATTCACACAGTGAATTAACCTTACCTTCTGCACTTTGGAT 1918
DB 542 CGGCGCTCTGTACGACATCGACCGCGCGGCGGACCTACCTGACCCACCTCTTGCACCTTCGGT 601
QY 1919 TCACAAGATCAGATATTTTAGCAATCACTCATAGAAGCGTGAGC-TGCAATGAATATG 1977
DB 602 ACACCCACATGGAGATCTTTAAGATCACCCACGCTTGGTCAACTTGCACCGCGGTGCTT 661
QY 1978 CGGAAAAACCGCGGTTTACTCTCAATTACCCGTCGATAGCCGCTGATTTTCAAACGTGC 2037
DB 562 GAGCGGACCGCGCTTACGCTTCACTACCCATCATTANTGTTGGCGGTTGAGGACGAC 721
QY 2038 AAGACTACGAGATGATCAAGG 2061
DB 722 ACCACGTTGGCGGGCTCCCAAG 745
RESULT 15
BH720365 713 bp DNA linear GSS 20-FEB-2002
LOCUS BOHSY82TF BO 2.3 KB Brassica oleracea genomic clone BOHSY82,
DEFINITION genomic survey sequence.
ACCESSION BH720365
VERSION BH720365.1 GI:18921108
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 713)
AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other GSSs: BOHSY82TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: ctown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: Tg
Class: sheared ends.
Location/Qualifiers
1. .713
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHSY82"
/clone_lib="BO_2.3_KB"
/notes="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOSt using BstXI linkers"
BASE COUNT 143 a 190 c 203 g 177 t
ORIGIN

```

Search completed: February 3, 2004, 00:23:53
Job time : 3287 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 09:08:17 ; Search time 21 Seconds
(without alignments)
3549.082 Million cell updates/sec

Title: US-09-806-767-2
Perfect score: 4018
Sequence: 1 MEKPFELCIIFLLFCSSSS.....NSHNLQVRSPISVTLXTN 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	4018	100.0	775	T00962	hypothetical prote
2	1742	43.4	754	A84473	probable serine pr
3	1679	41.8	775	T07172	subtilisin-like pr
4	1666.5	41.5	757	JC7519	subtilisin-like se
5	1640.5	40.8	746	S52770	subtilisin-like pr
6	1611	40.1	766	T07171	subtilisin-like pr
7	1589	39.5	784	T05768	subtilisin-like pr
8	1572.5	39.1	754	T06577	subtilisin-like pr
9	1562	38.9	756	G86150	F22M8.3 protein -
10	1542	38.4	754	T06579	subtilisin-like pr
11	1534.5	38.2	745	T07184	subtilisin-like pr
12	1513.5	37.7	747	T06580	subtilisin-like pr
13	1490	37.1	745	JC6119	subtilisin-like pr
14	1445.5	36.0	758	T51335	subtilisin-like pr
15	1282	31.9	761	S52769	subtilisin-like pr
16	1272	31.7	774	A86454	hypothetical prote
17	1271.5	31.6	756	T04187	subtilisin-like pr
18	1236	30.8	777	C86454	hypothetical prote
19	1221	30.4	766	T48389	cucumis-like pro
20	1220.5	30.4	761	T07169	subtilisin-like pr
21	1219.5	30.4	779	C86335	hypothetical prote
22	1216.5	30.3	763	C86454	hypothetical prote
23	1215.5	30.3	775	T04189	subtilisin-like pr
24	1213	30.2	769	D86335	T20H2.6 protein -
25	1201	29.9	736	JC7518	subtilisin-like se
26	1196	29.8	803	T04190	subtilisin-like pr
27	1190.5	29.6	747	T04188	subtilisin-like pr
28	1188.5	29.6	779	T07170	subtilisin-like pr
29	1187.5	29.6	706	T01351	subtilisin-like pr

30	1180.5	29.4	736	2	T12963	subtilisin homolog
31	1176	29.3	739	2	T12964	subtilisin homolog
32	1173	29.2	765	2	T04186	subtilisin-like pr
33	1172.5	29.2	731	1	A55800	cucumis (EC 3.4.
34	1145.5	28.5	746	2	T06017	subtilisin-like pr
35	1143	28.4	753	2	B96687	subtilisin-like pr
36	1134.5	28.2	758	2	A96887	hypothetical prote
37	1117.5	27.8	779	2	T14845	antifreeze-like pr
38	1117	27.8	772	2	T05838	subtilisin-like pr
39	1101.5	27.4	755	2	T48553	subtilisin-like pr
40	1076.5	26.8	718	2	T05840	subtilisin-like pr
41	1070	26.6	769	2	T05939	subtilisin-like pr
42	1067	26.6	1736	2	T05174	hypothetical prote
43	989	24.6	687	2	D85165	cucumis (importe
44	985	24.5	734	2	D86454	F9L11.14 F9L11.14
45	907.5	22.6	815	2	T00538	probable serine pr

ALIGNMENTS

RESULT 1

T00962
hypothetical protein F20D22.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C:Accession: T00962
R:Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu, G.; Oji, K.; Feng, J.; Kim, C.; Kurtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC F20D22 complete sequence.
A:Reference number: Z14214
A:Accession: T00962
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-775 <VYS>
A:Cross-references: EMBL:AC002411; NID:G2570223; PID:G3142298; GSPDB:GN00059; ATSP:F20L
C:Genetics:
A:Gene: ATSP:F20D22.12
A:Map position: 1
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homolog

Query Match	100.0%	Score 4018	DB 2	Length 775
Best Local Similarity	100.0%	Pred. No. 2.5e-270		
Matches	775	Conservative	0	Mismatches
Indels	0	Gaps	0	
QY	1	MEKPFELCIIFLLFCSSSEILQKQYIVQLHPNSTAKTFASKFDWHLSEFLQEAVLGV	60	
Db	1	MEKPFELCIIFLLFCSSSEILQKQYIVQLHPNSTAKTFASKFDWHLSEFLQEAVLGV	60	
QY	61	EEEEEPSSRLLYSGAIGFAAQLTSEAEILRYSPEVAVRPHVLQVQITTSYKFL	120	
Db	61	EEEEEPSSRLLYSGAIGFAAQLTSEAEILRYSPEVAVRPHVLQVQITTSYKFL	120	
QY	121	GLDGFNGSWGSKRFGQGTIIIGVLDGVMPSPFDGTGMPSPRKKWICQEGSFSS	180	
Db	121	GLDGFNGSWGSKRFGQGTIIIGVLDGVMPSPFDGTGMPSPRKKWICQEGSFSS	180	
QY	181	SSCNKRLIGARFFIRGHRVANSPEESPMPREYISARDSTGCHGTHTASTVGGSSVSMANV	240	
Db	181	SSCNKRLIGARFFIRGHRVANSPEESPMPREYISARDSTGCHGTHTASTVGGSSVSMANV	240	
QY	241	LGNGAGVARGMAPGAHIAVYKVCWFGNGCYSSDILAAIDVAIQDVLSLGGFPPIPLY	300	
Db	241	LGNGAGVARGMAPGAHIAVYKVCWFGNGCYSSDILAAIDVAIQDVLSLGGFPPIPLY	300	
QY	301	DDTATGCTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTIDRRFPAPVVRANG	360	
Db	301	DDTATGCTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTIDRRFPAPVVRANG	360	
QY	361	KLLYGESLYPGKGNAGREVEVIYVGGXGSEFLRGLSPREIRGKWCIDRGVNGR	420	
Db	361	KLLYGESLYPGKGNAGREVEVIYVGGXGSEFLRGLSPREIRGKWCIDRGVNGR	420	

```

Qy 421 SEGEAVKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYVNAVTKPKAR 480
Db 421 SEGEAVKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYVNAVTKPKAR 480
Qy 481 IFPGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMAIPGVNIIAAMPQNLGPTGLPYDS 540
Db 481 IFPGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMAIPGVNIIAAMPQNLGPTGLPYDS 540
Qy 541 RRVNFTVMSGTSMSCHVSGITATIRSAYPNNWSPAAIKSALMTTADLDYRQKAKIDGNK 600
Db 541 RRVNFTVMSGTSMSCHVSGITATIRSAYPNNWSPAAIKSALMTTADLDYRQKAKIDGNK 600
Qy 601 PAGVFAIGAGHVNPOKALNPLGVNIQPDYITLCTLGFTSDILAIHKNVSCNGILR 660
Db 601 PAGVFAIGAGHVNPOKALNPLGVNIQPDYITLCTLGFTSDILAIHKNVSCNGILR 660
Qy 661 KNPFGSLNPSIAVIFKRGKTTMTIRRTVNVGSPNSIYVNVKAPGEGIKVIVNPKRLVF 720
Db 661 KNPFGSLNPSIAVIFKRGKTTMTIRRTVNVGSPNSIYVNVKAPGEGIKVIVNPKRLVF 720
Qy 721 KHVDTLSYRVFVWLKKORGGKVASFAOQLTWNHNLQVRSPISVTUKTN 775
Db 721 KHVDTLSYRVFVWLKKORGGKVASFAOQLTWNHNLQVRSPISVTUKTN 775

RESULT 2
A84473
Probable serine proteinase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: A84473
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Euss, D.; Nierman, W.C.; White, O.; Eiser, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; PMID:20083487; PMID:10617197
A:Accession: A84473
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-754 <SPO>
A:Cross-references: GB:AE002093; NID:G4006827; PIDN:AAC95169.1; GSPDB:GN00139
C:Gene: At2g05920
A:Map position: 2
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology

Query Match 43.4%; Score 1742; DB 2; Length 754;
Best Local Similarity 48.4%; Pred. No. 1.3e-112;
Matches 370; Conservative 116; Mismatches 237; Indels 40; Gaps 16;

Qy 11 IFLLFSSSEILQKQTYIVQLHPNSETAKTRASKPDWHLFLQEAFLGVEEERPSR 70
Db 18 LFLDLHTA-----KTYIIIRVN-HSKPESFTHDWTYSQNS-----ESS 59

Qy 71 LLYSGSAIGRAQLTESEAE-ILRYSPEVAVRDPHVLQVQTYTSYKFLGDLG-FGNS 128
Db 60 LLYTYTTFHGFSAYLSTEDSLSSNSILDFEDPLYLTHTRTPFLGLNSEFGVH 119

Qy 129 GVWSKSRFGQGTIIIGVLDGWVPSRFDGTGMPSPRKPWKGCQGESESSSNCKLI 188
Db 120 DLGSSS---NGVIIIGVLDGWVPSRFDGTGMPSPRKPWKGCESGSDPSKLNCKLI 176

Qy 189 GARFIRGHRVANSPSPNPRYISARDSTGHGTASTVGGSSVSMANVLGNGVYA 248
Db 177 GARSFKGFQVAGSGGFSK--RESVSPRDVDGHTHTSTTAAGASVARNASFLYGAGTA 234

Qy 249 RGMAPGAHIAVYKVCWFGYSSDILAAIDVAIDKDYVLSLSLGGPPIPLYDDTTAIGT 308
Db 235 RGMATRAVATYKVCWSTGCGFSGDILAMRAILDGVDVLSLGGGSAFYRDTTIGA 294

Qy 309 FRAMERGISVICAAGNNGPIESSVANTAPWVSTTGAGTLDRFPFVAVRVLANGKLYGESL 368

```

```

Db 295 FSAVERGVFVSCSAGNSGPTASVANVAPWMTVAGTLDLDRFPFANLNGKRLTGVS 354
Qy 369 YPGKGIKAGREVETVYTGDKGSFCLRGSLPREIIRGKMWICDRGVNGRSEKGEAVK 428
Db 355 YSGVGM--GTPLELVNKGNSSSNLCPLGSLDSSIVRGKIVCDRGVNARVEKAVR 412
Qy 429 EAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYVNAVTKPKARIIFOGTVI 488
Db 413 DAGGLGMIMANTAAAGBELVADSHLLPAITAVGKTKDGLLREYVKSDSKPTALLVFKGTVL 472
Qy 489 GRSRAPEVAQFSARGPSLANPSILKPDMTAPGVNIIAAMPQNLGPTGLPYDSRVNFTUM 548
Db 473 DVKPSPVAAAFSSRGENTVTEILKPDVIGPGVNIAGWSDAIGPTGLDKDSRTQFNM 532
Qy 549 SGTSMSCPHVSGITATIRSAYPNNWSPAAIKSALMTTADLDYRQKAKIDG--NKPAGVPA 606
Db 533 SGTSMSCPHVSGITATIRSAYPNNWSPAAIKSALMTTAYVLDNTNAPLHDAADNSLNPYA 592
Qy 607 IGAGHVNPOKALNPLGVNIQPDYITLCTLGFTSDILAIHKNVSCNGILRKNPGF 665
Db 593 HSGSHVDPOKALSGPLVYDITSTEYIRFLCSLDYTVDHVAIVKRPVSNCKKF--SDPG-- 650
Qy 666 SLNYPSTAVIFKRGKTTMTIRRTVNVGSPNSIYVNVKAPGEGIKVIVNPKRLVFKHVDQ 725
Db 651 QLNYPSPSVLF--GSKRVVRYTREVTNVGAASSVYKVTNAGAPSVGIVKPSKLSFKSVGE 709

Qy 726 TLSYRVFVWLKKORGGKVASFAO-QQLTWNHNLQVRSPIS 769
Db 710 KKRYTVTFVSKK---GVSMTNKAEFGSITWSNPQH---EVRSPVA 748

RESULT 3
T07172
subtilisin-like proteinase (EC 3.4.21.-) 2 - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C:Accession: T07172
R:Meichtry, J.; Amrhein, N.; Schaller, A.
submitted to the EMBL Data Library, May 1998
A:Description: The gene family of subtilisin-like proteases in tomato.
A:Reference number: 215771
A:Accession: T07172
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-775 <MEI>
A:Cross-references: EMBL:AJ006379; PIDN:CAA07000.1
A:Experimental source: cultivar VFw8
C:Genetics:
A:Gene: sbt2
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
C:Keywords: hydrolase; serine proteinase

Query Match 41.8%; Score 1679; DB 2; Length 775;
Best Local Similarity 46.0%; Pred. No. 3.2e-108;
Matches 359; Conservative 125; Mismatches 261; Indels 36; Gaps 16;

Qy 6 FFLCIT---IFLLFSSSEILQKQTYIVQLHPNSETAK--TFASKFDWHLFLQEAFLGV 60
Db 11 FFVSVCILAINLAKCSPNT---KTYIIQM--DKWAKPDVFDVHVQWYSSSLKSVLPST 63

Qy 61 EEEEE--EPPSRLLYSYGSALTEGFAAQLTESEAILRYSPEVAVRDPHVLQVQTYTSYK 118
Db 64 TEVENTGDGEIRILYSYQTAFFHVAQAQLSEEVKQLQERNGVLAVFFPEIKYQLHTTSPL 123

Qy 119 FLGLDGFNSGVWSKSRFGQGTIIIGVLDGWVPSRFDGTGMPSPRKPWKGCQGESESF 178
Db 124 FLGLDREDSKILADRLSDHNVIIGVLDGTGWPSRFDGNTSVSHWKGVCETGRGF 183

Qy 179 SSSSNCKRLICARFFIRGHRVANSPSPNPRYISARDSTGHGTHTASTVGGSSYSMA 238
Db 184 EXHCKSKIVGARVFRGFEAAAG---KINERGFKSARDQDGHGHTAGTVAGSVVRGA 240

```

```

QY 239 NVLGNAGVARGMAPGAHIAVYKVCWFNGCYSSDILAAIDVAIDOKVDVLSLSLGGPPIP 298
DB 241 NLLGAYGTARGMAPGAHIAVYKVCWFNGCYSSDILAAIDVAIDOKVDVLSLSLGGVSS 300
QY 299 LYDDTIACTFRAMERGISVCAAGNNGPIESSVANTAPWSTIGACTLDRRPPAVVRLA 358
DB 301 YNRSLSTAAEGAMEKGVFVSCSAGNGGPDPISTVNSPWITTVGASTMDRDPATVELG 360
QY 359 NGKLLYGBSLYPGKIRKAGREVEVYVYTGDKG---SEFCRLGSLPREIRGQWVICDR 415
DB 361 TGKIVTGASLYKGRMNLSTQKQYPLIYLGSSNLSMPSSSLCDGLDGLKASVAGKIVICDR 420
QY 416 GNGRSEKGEAVKEAGGVAMILANTEINOEDSDIVHLLPATLIGYTESVLLKAYVNAVY 475
DB 421 GISPREVQGVYKGEAGGVCMILTNTAANGEELVADSHLLPAVAVGEREGRAIKDYA-AGR 479
QY 476 KPKARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPMIAPGVNIIAAMPONLQPTG 535
DB 480 SATATLRLGLTKGIRPSVVAARSRGPNFLSLLEILKPMVAPGVNIIKAGTALGPSS 539
QY 536 LPYDSRRVFTVMSGTSMSCPHVSGITALLIRSAYPNWSPPAAIKSALMTTADLYDRQKAI 595
DB 540 LPIDQRRNFNLSTGSMSCPHVSGIAALLKARHPDWSPPAAIKSALMTTAVVHDNTYKSL 599
QY 596 KGNK--PAGVFAIGAGHVNPKAINPGLVNIQPVYIYVLCGLGTRSDILAI-THKN 652
DB 600 KDASVTPTSDYDHGAGHVNPRKAVDPLGIYDIGAQDYFEFLCTQELSPSQLVFGKFSN 659
QY 653 VSCGILRKNGPFSINPSIAVIFKRGKTTMIT--RRTVNGSPNSIYSVNYKAPGK 710
DB 660 RTCHSL-ANPG-DLNPAPISAVPEKTKLSMLTLHRTVTVNGSPISYHVVVSAFAGV 717
QY 711 VIVNPKRLVFKHVDQTLRYVWFVLLKKNGRGKVASFAQGOLTVWNSHNLQVRSPISV 770
DB 718 VKVEPERLNFSTKQKLSYKTKTVSRQ---KAPF--GSLIKWKG---THKVRSPIAI 769
QY 771 T 771
DB 770 T 770

RESULT 4
JC7519
subtilisin-like serine proteinase (EC 3.4.21.-) - Arabidopsis thaliana
N:Alternate names: subtilase
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
C:Accession: JC7519
R:Yamagata, H.; Usugi, M.; Saka, K.; Iwasaki, T.; Aizono, Y.
Biosci. Biotechnol. Biochem. 64, 1947-1957, 2000
A>Title: Molecular cloning and characterization of a cDNA and a gene for subtilisin-like
A:Reference number: JC7518; MUID:20506592; PMID:11055401
A:Accession: JC7519
A:Molecule type: DNA
A:Residues: 1-757 <YAM>
A:Cross-references: GB:AF065639
C:Comment: This enzyme is organ-specific and is involved in the specific proteolytic eye
C:Genetics:
A:Gene: asp48
C:Superfamily: subtilisin-like proteinase ag12; subtilisin homology
C:Keywords: hydrolase; serine proteinase

Query Match 41.5%; Score 1666.5; DB 2; Length 757;
Best Local Similarity 46.4%; Pred. NO. 2.3e-107;
Matches 362; Conservative 122; Mismatches 246; Indels 51; Gaps 18;

QY 6 PFLCIFLPCSSSEILQKQYIVQLHPNSETAKT-PASKFDWHLFLQEAVLGVEEE 64
DB 11 PFL-LLCGFCRVSSSSDQQTIVHM-----AKQMPSSFDLHSNWDSSLRISD-- 61
QY 65 EEPSSRLLYSGAIEGPAALTESEABILRYSPWEVAVRDPHVLQVQTTYSYKFLGLDG 124
DB 62 ---SALLIYTYENALHGSTRLTQEDADSLMTQGVISVLPEHRYELHTTTPFLGLDE 118

```

```

QY 125 -----FGNSGVWKSRRFGGTLIGVLDTGVPWSPSPDDTGMPSTPRKWKGLCQEGESFS 179
DB 119 HTADLFPAGSYSS-----DVVVGVDLTGVWPESXSYSDEGFGPIPSWKKGCEAGTNT 172
QY 180 SSSCNRLKILGARFFIRGHRVANSP--EESPNNPREVISARDSTGHTHTASTVGGSSVMA 238
DB 173 ASLCNRKILGARFPARGVESTMGPIDES-----KESRSPRDDDGHTHTSSAAGSVVGA 228
QY 239 NVLGNAGVARGMAPGAHIAVYKVCWFNGCYSSDILAAIDVAIDOKVDVLSLSLGGPPIP 298
DB 229 SLGVSAGTARGMAPRARVAVYKVCWLGCGCFSSDILAAIDKAIADNVANVLSLGGMSD 288
QY 299 LYDDTIACTFRAMERGISVCAAGNNGPIESSVANTAPWSTIGACTLDRRPPAVVRLA 358
DB 289 YTRDGVAIAGFAAMERGIIVSCSAGNAGPSSSLSNVAPWITTVGAGTLDLDRFPALALIG 348
QY 359 NGKLLYGBSLYPGKIRKAGREVEVYVYTGDKG--EFCLGSLPREIRGQWVICDRG 416
DB 349 NGKNFTGSLFKGEALPD--KLLPFIYAGNASNATNGMLCMTGLIPEKVKGIWMCDRG 406
QY 417 VNGRSEKGEAVKEAGGVAMILANTEINOEDSDIVHLLPATLIGYTESVLLKAYVNAVYK 476
DB 407 INARVQKGDVYKAGGVNLANAANGELVADAHLLPATVGEKAGDIIRHYVTTDPN 466
QY 477 PKARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPMIAPGVNIIAAMPONLQPTGL 536
DB 467 PTASISILGTVVGVKPSVVAARFSSRGPNSTIPNLKPDLLIAPGVNIIAAMPONLQPTGL 526
QY 537 PYDSRRVFTVMSGTSMSCPHVSGITALLIRSAYPNWSPPAAIKSALMTTADLYDRQKAIK 596
DB 527 ASDSRVEFNIISGTSMSCPHVSGLAALLKSVHPEWSPAAIRSAALMTTAYKTYKDGKPLL 586
QY 597 D--GNKPAGVFAIGAGHVNPKAINPGLVNIQPVYIYVLCGLGTRSDILAIHKNVS 654
DB 587 DIATGKPSPPDHGAGHVSPTTATNPGLIYDUTTEDYLGFLCALNYTSPQIRSVRRNT 646
QY 655 CNGILRKNGPFS--LNPYSIAVIFKRGKTTMIT--RRTVNGSPNSIYSVNYKA-PEGIK 710
DB 647 CD----PSKSYSVADLNPYSPFAVVD--GVGAKYKTRTVTSVGGAGT--YVKYVTSSETGVK 700
QY 711 VIVNPKRLVFKHVDQTLRYVWFVLLKKNGRGKVASFAQGOLTVWNSHNLQVRSPISV 770
DB 701 ISVEPAVLNFKEANERKSYTVTFTVDSKPSG---SNSFGSIEWSDGKHV---VGSVAI 754
QY 771 T 771
DB 755 S 755

RESULT 5
S52770
subtilisin-like proteinase (EC 3.4.21.-), nodule-specific - Arabidopsis thaliana (frag
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 22-Jun-1999
C:Accession: S52770
R:Ribeiro, A.; Akkermans, A.D.L.; van Kammen, A.; Bisseling, T.; Pawlowski, K.
submitted to the EMBL Data Library, March 1995
A:Description: A nodule-specific gene encoding a subtilisin-like protease is involved :
A:Reference number: S52769
A:Accession: S52770
A:Molecule type: mRNA
A:Residues: 1-746 <RIB>
A:Cross-references: EMBL:X85974; NID:g757533; PIDN:CAAS9963.1; PID:g757534
C:Function:
A:Description: Involved in early stages of actinorhizal nodule development
C:Superfamily: subtilisin-like proteinase ag12; subtilisin homology
C:Keywords: hydrolase; serine proteinase

Query Match 40.8%; Score 1640.5; DB 2; Length 746;
Best Local Similarity 46.0%; Pred. NO. 1.4e-105;
Matches 359; Conservative 122; Mismatches 248; Indels 51; Gaps 19;

```

[illegible]

RESULT 6

T07171
subtilisin-like proteinase (EC 3.4.21.-) 1 - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C:Accession: T07171
R:McIlhtry, J.; Amrhein, N.; Schaller, A.
submitted to the EMBL Data Library, May 1998
A:Description: The gene family of subtilisin-like proteases in tomato.
A:Reference number: Z15771

A;Accession: T07171
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-766 <MEI>
A;Cross-references: EMBL:A006378; PIDDN:CAI06999.1
A;Experimental source: cultivar VFw8

C;Generics:
A;Gene: sbt1
C;Superfamily: subtilisin-like proteinase ag12; subtilisin homology
C;Keywords: hydrolase; serine proteinase

Query Match	40.1%;	Score 1611;	DB 2;	Length 766;
Best Local Similarity	43.3%;	Pred. No. 1.6e-103;		
Matches	339;	Conservative 140;	Mismatches 250;	Indels 54; Gaps 17;
Qy	8	LCIIFLLFCSSSEILQKQYIYVQLHPNSSTAKTFASKEDWHLISFLQEAVLGVVEE	EEP	67
Db	13	MVVLHFVYDARQN--QKTYIITHMDKFNMPA-----DFDDHTQWYDSSLUKSVK	----	60
Qy	68	SSRLISYSGAIEGFAAQLTSEAEILRYSPVVAVRPDHLVQVQTYYSYKFLGDGF	GN	127
Db	61	SANWLYTNSVINGYSTQLTADAKALAQPGILLVHEEVIYELHTRSTFTFLG	EGRES	120
Qy	128	SGVMSKSRFQCGTIIGVLDYGVMPESPFDTCMPISIPRKWKIGCQSGESFSS	SNRKL	187
Db	121	RSFPQTEARSEVLIGVLDYGVMPESKSFDDTGLQVPAWKGKQCTQKXNF	DASSCN	180
Qy	188	IGARFFTRGHRVA--NSPEESPNNPREVISAEDETGHGHTHTASTVGGSSVS	MANVLNGAG	246
Db	181	IGARFFSQGYEAAFGAIDET---IESKSPRDDEGHGTHTATTAAAGSVW	TGASLLGYATG	236
Qy	247	VARGMAPGAHIAVYKVCWFNGCYSSDILAAIDVAIOQVDVLSLSCGFFPL	YDDYIAI	306
Db	237	TAGMASHARVAAYKVCWTGCGFSSDILAGMDQAVIDGNVLSLGGTIS	DYHERDVAI	296
Qy	307	GTFRAMERGISVICAAGNNGPIESSVANTAPWYSTIGAGTLDRRFP	AVRVLANKL	366
Db	297	GAFSAASQGI FVSCSAGNGGSPSOTLSNAPWITTVGAGTMDREF	PAYIGIGNCK	356
Qy	367	SLYPGKGIKNAGREVEVIYTGDKGS--BFCURGSLPREIRGQWVIC	RGVNGREK	424
Db	357	SLYSGKALPSS--VMPLVYAGNVQSNGNLCTSGSLIPEKVAGKI	VVCDRGMN	414
Qy	425	ENKKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYV	NATVKPKARI	484
Db	415	LVVKDAGGIGMILANTDTYGDELVDADHLLPTAAVGQTAGNLIKQY	IASNENPI	474
Qy	485	GTVIGRAREPAVQAFSARGPSLANPSILKPDMIAPGVNIIAAMPQ	NLCPTGLP	544
Db	475	GTKLVQPSPVWAAPFSRGGNPITPDVLKDLLAPGVNILLAGWT	KVGPTGLQ	534
Qy	545	FTVMSGTSMCPHVSGITALIRSAYPNWSPAATKALMTTADLY	VDROGKA	599
Db	535	FNIIISGTSMSCPVSGLAALLKAAHPWSPAAIRSAALMTTSY	TYKNGKTI	594
Qy	600	KPAGVFAIGAGHUNPQKAINPLVYNTPQVDYITLCTLGTFR	SDILLATH	659
Db	595	TP---FDYGAGHVNPTAAVSPGLVYDLTVDPDYNFLCALD	YSPMKVI	648
Qy	660	RKHPGF---SINYPISIAVIFKRG-----KTTEMLTRRV	TNVGSPNS	706
Db	649	-ENKEYRVDLUNYPSFSIPMETANGEHADSSPTVTRYTR	TLTNVGNP	706
Qy	707	EGIKVIVNPKRLVPKHVDQTLISYRVFVLKXKNRGKGV	SPAQGLT	766
Db	707	QDVKILVEPQTLFPRKNEKTYITVTTATSKPSG--TT	SFA--R	759
Qy	767	PIS	769	
Db	760	PIA	762	

RESULT 7

subtilisin-like proteinase (EC 3.4.21.-) - Arabidopsis thaliana
T05768
N:Alternate names: protein M4E13.40
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change
A:Accession: T05768
R:Bevan, M.; Purnelle, B.; Boutry, M.; Goffeau, A.; Hoheisel, J
submitted to the Protein Sequence database, April 1998
A:Reference number: Z15451
A:Accession: T05768

[illegible]

Qy	707	E G I K V I N P R L V F H Q V D T L S Y X W F V L K C K N R G K V A F Q A G O L T W V N S H N L M Q R V R S	766
	:	: : : : :	:
	:	: : : : :	:
	:	: : : : :	:
	:	: : : : :	:
Db	686	K G V V V K Y P K R L P S E L K Q L T Y Q V T F - - S K R T N S S K G V P - E G F L K W - N E N K - - Y S V R S	739
	:	: : : : :	:
	:	: : : : :	:
	:	: : : : :	:
	:	: : : : :	:
Qy	767	P I S V	770

```

RESULT 8
T06577
subtilisin-like proteinase (EC 3.4.21.-) - tomato
N:Alternate names: proteinase p69c
C:Species: Lycopersicon esculentum (tomato)
C:Date: 23-Apr-1999 #sequence revision 23-Apr-1999 #text
change 20-Jun-2000

```



```

Db      740 PIAV 743
||||
RESULT 9
G86150
F22M8.3 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C:Accession: G86150
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86150
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-756 <STO>
A:Cross-references: GB:AE005172; NID:G8570441; PIDN:AAF76468.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology

Query Match      38.9%; Score 1562; DB 2; Length 756;
Best Local Similarity 44.0%; Pred. No. 4e-100;
Matches 341; Conservative 130; Mismatches 268; Indels 36; Gaps 18;

Qy      7 FLCIFLLFCSSSEILQKTYIVQLHPNSETAKTFASKFDHLSFLQEAVALGVEEBEE 66
Db      6 FIVFFLLFASNVUS-SRKTYVHTVTTTS-TKHIVTSLFN---SLQTENI---NDDDF 56

Qy      67 PSSRLLYSGAIEGFAAQTESEAEILRYSPDEVAVRPHVLQVOTYSYKFLGDGFG 126
Db      57 SLPEIHYIENAMSGFATLTDQLDTVKNTKGFISAYPDDELLSLHTYSHEFLGLE-FG 115

Qy      127 NSGVWSKSRGQGTIIIGLVGTWPSPSPDDTGMPSPRKKWGIQCEGSPSSSCNRK 186
Db      116 -IGLWNETSLSSDVIIGLVGTGISPEHVSPRDTHTMTVPVRWRCGDEGNFSSSCNRK 174

Qy      187 LIGARFTRGRHRVANSPEESPNPREYISARDSTGHGTHASTVGGSSVMANVLNGAG 246
Db      175 IIGASAFYKY---ESIVGKINETTDFSTRDAQGHGTHASTAAGDIVPKANYFGQAKG 231

Qy      247 VARGWAPGAHTAVYKVCNFGCYSSDILAAIDVAIQKVDVLSLGLGFPPIPLYDTTIAI 306
Db      232 LASGWRFTSRIAAYKACWALGCASTDVIADRAILDGVDVLSLGLGSSRPFPVDPPIAI 291

Qy      307 GTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTLDRPFPVAVELANGKLLYGE 366
Db      292 AGFGAMQKNIFVSCAGNSGFTASTVNGAPWLTVAASYTDRTPFAIVRIGNRKSIVGS 351

Qy      367 SLYPKGKKNAGREVIYVYGGKGFCLRGSLPREETRGKAVICDRGVNCRSEKGEA 426
Db      352 SLYKGSLSKLN--PLAFNRTAGEESGAVFCIRDSLKRELVEGKIVICLRASGRTAKGEE 409

Qy      427 VKEAGGVAMILANTEINEEDSIDVHLLPATLIGYTESVLLKAYVNAVTPFKARIFFGGT 486
Db      410 VKRSGGAALLVSTEAEGEELLADPHVLPVAVSLGFSGDKTLNLYLAGANATASVRFGT 469

Qy      487 VIGSRAPVQAQFARGPSLANPSILKPDMTAPGNIIAWPNQLGPTGLPYDSRRNVFT 546
Db      470 AYG-ATAPMVAAFSSRGVSAGPEIAKPDIAAPGLNLAGSPSPSLLRSPPRRYQFN 528

Qy      547 VMSGTSMSCPHVSGITALIRSAFYNWSPAAIKSALMTTADLYDQGKAID-----GNKP 601
Db      529 IISGTSMACPHISGIAALIKSVHGDWSPAMIKSAIMTITARNRPIGDRGAAGAES 588

```

```

Qy      602 AGVFAIGAGHVNPOKAINPGLVNIQFVYIYVLTCLGTFRSDILAITHQVSC--NGIL 559
Db      589 ATAFAGAGNVDPTRAVDPLGVYDTSTVDYLNLYCSLNTSETRIILFSGTNYTCASNAV 648

Qy      660 RKNPGPSLNYPSIAVIFKRGKTTMI--TRRTNVGSPNSIYSVNVKAPGEGIVNPKR 717
Db      649 -LSFG-DLNPSPFAVNLVNGANLKTVRYKRTVTNVGSPICEYMHVVEEPKGVPEPKV 706

Qy      718 LVFKHVDOTLSYRWVFLVKKKNGGKVASPAQCLTWV-NSHNLMOVRSPISVT 771
Db      707 LKFOKAEERLSYTVYD-AEASNSSSSP--GVLAWICDKYN---VRSPIAVT 754

RESULT 10
T06579
subtilisin-like proteinase (EC 3.4.21.-) p69e - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C:Accession: T06579
R:Meichtry, J.; Amrhein, N.; Schaller, A.
submitted to the EMBL Data Library, May 1998
A:Description: The gene family of subtilisin-like proteases in tomato.
A:Reference number: Z15771
A:Accession: T06579
A>Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-754 <MEI>
A:Cross-references: EMBL:AJ005172; PIDN:CAA06413.1
A:Experimental source: cultivar VFWM
C:Genetics:
A:Gene: p69e
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
C:Keywords: hydrolase; serine proteinase

Query Match      38.4%; Score 1542; DB 2; Length 754;
Best Local Similarity 43.4%; Pred. No. 9.7e-99;
Matches 341; Conservative 119; Mismatches 261; Indels 64; Gaps 19;

Qy      6 FLCIFLLFCSSSEILQK--QTYIVQLH-PNS--ETAKTFASKFDHLSFLQEAVALGV 60
Db      3 FLKILLVLIFCSFPWPTIQSGLEIYIVHVESPELSLSTQSFDTLDSYYISFLPETTSAI 62

Qy      61 EEEEEPSSRLLYSGVNAIEGFAAQTESEAEILRYSPDEVAVRPHVLQVOTYSYKFL 120
Db      63 SSSGNEEARWTIYSHNVMTGFAARLTASHVENEKRGFVSAQKQILSLDTHTSFL 122

Qy      121 GLDGFSGSVWSKSRFGQGTIIIGLVGTWPSPSPDDTGMPSPRKKWGIQCEGSEFSS 180
Db      123 GLQ--QNWGVWKDSNYGKGVIIIGVLDGILPDHPSFDVGMPPPPAKWKGYC---ESNFT 177

Qy      181 SSCNRKLI GARFFTRGRHRVANSPEESPNPREYISARDSTGHGTHASTVGGSSVMANV 240
Db      178 NKCNRKLI GARSYHLNG-----SPIDGDGHGTHASTAAGAFVKGANV 221

Qy      241 LONGAGVARGVAPGAHTAVYKVCNFG--GCYSSDILAAIDVAIQKVDVLSLGLGFPIL 299
Db      222 YGNANGTAVGAPLAHTAVYKVCSSDGGCSDSILAAIMDSALDGDVILSISIGGSPNSL 281

Qy      300 YDDTIAIGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTLDRPFPVAVLAN 359
Db      282 YDDPIALGAYSATARGVFCVSCAGNRPGLLASVGNAPWILTGVGASTLDRKIKATVGLGN 341

Qy      360 GKLLYGESLYPKK-----GKNAGREVIYVYGGKGFCLRGSLPREETRG 408
Db      342 GEEFEGESAYRPQTSNSTFTFLDAAKHAKDPSETPY-----CRPGSLTDPVIRG 391

Qy      409 KMWIC--DRGVNCRSEKGEAEVKEAGGVAMILANTEINEEDSIDVHLLPATLIGYTESVL 466
Db      392 KTVLCACGGVSS--VDKGVKWDAGGVGVINPNSQYGVTKSADAHVLPALDVSADGTR 450

Qy      467 LKAYNAVTPFKARIIFGTVIGRSPAFAVQAQFARGPSLANPSILKPDMIAPGVNIIAA 526

```

Db 451 IRAYTNSILNPVATITFOGTHIGDENAPIVAAPSSRGPNTPASGILKPDIIIGPGVNIILAA 510
 QY 527 WPNLPGTGLPYDSRRVNTVMGTSMSCPHVSIGITALIRSYPNWSPAAIKSALMTTAD 586
 Db 511 WPTSVDCN-----KNTKSTENIIISGTSMSCPHLSGVAALLKSHPDWSPAVIKSAIMTTAD 566
 QY 587 LYDRQGAIAKDGK-KEAGVPAIGAGHVNPKAINPGLVNIQPDVYITVLTGFTSRDI 645
 Db 567 TNLASSPIIDELKSLADIIAIGAGHVNPSRANDPGLVYDTFPEDVLPYLCGLNLTNSQV 626
 QY 646 LAITHKNVSCNGILRNKPNPFLNYPSTIAVIFKRGKTKTEMITRRVTNVGSPNSIYSVNVKA 705
 Db 627 GKLLKKNVCSV-ESIPAEQLNYPFC-ISRLGSPOTFTPTVTVNVGDAKSYTVQIAS 684
 QY 706 PEGIKVINPKRIVKHVDOTLSYRVFVVKKNRGKGVASPAQGLTWNHSHLMQVR 765
 Db 685 PKGVVVKPRKLIFFSELKQKLYQVTF--SKRTNSKSGVF-EGFLKW-NSNK--YSVR 738
 QY 766 SPISV 770
 Db 739 SPIAV 743

RESULT 11
 T07184
 subtilisin-like proteinase (EC 3.4.21.-) precursor P69B, pathogenesis-related - tomato
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 21-Jul-2000
 C:Accession: T07184
 R:Toranzo, P.; Conejero, V.; Vera, P.
 J. Biol. Chem. 272, 14412-14419, 1997
 A:Title: Identification of a new pathogen-induced member of the subtilisin-like processin
 A:Reference number: Z15980; MUID:97306355; PMID:9162080
 A:Accession: T07184
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-745 <TOR>
 A:Cross-references: EMBL:Y10149; NID:G2230958; PIDN:CAA71234.1; PID:G2230959
 A:Experimental source: cultivar Rutgers; leaf
 C:Genetics:
 A:Gene: P69B
 C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
 C:Keywords: hydrolase; serine proteinase
 F:1-115/Domain: signal sequence and amino-terminal propeptide #status predicted <PRO>
 F:115-745/Product: subtilisin-like proteinase #status predicted <NAT>

Query Match 38.2%; Score 1534.5; DB 2; Length 745;
 Best Local Similarity 43.3%; Pred. No. 3.2e-98;
 Matches 339; Conservative 109; Mismatches 266; Indels 69; Gaps 17;

QY 10 IIFLLFCCSSSEILQK--QTYIVQLH-PNS--ETAKTFASKPDWHLSPFLQEAVLGVVEEE 64
 Db 7 LLVFIFCSFQWPTIQSNLETVIVHVSPELVTTQSLLDLGSYYLSFLPKTATTISSG 66
 QY 65 EEPSSRLLYSGAIEGFAAQLTESAEILRYSPEVVAVRPHVLQVQTYTSYKFLGLDG 124
 Db 67 NEENATWLYSYHNVMGFAARLTAQVKENEKKGHFVSACKQRIILSHLTHTPSFLGLQ- 125
 QY 125 FNGSGWWSKRFQGTIIIGVLDTGVPSPSPDGTGMPSPRKGICQEGESFSSSCN 184
 Db 126 -QNGVWKNDSYKGVIGVLDGILPDHPSFDVGMGPPPAKWKVC--ESNFTKCN 181
 QY 185 KLIIGARFFTRGHRVANSPEESPMPREYISARDSTGHGHTHTASTVGGSSVSMANVLNG 244
 Db 182 NKLIIGARSYQLNG-----SPIDSGHGTHTASTAAGAFVKGANYGNA 225
 QY 245 AGVARGMAPGAHTAVYKVCWFGNGCYSDILAAIDVAIQDKVDVLSLGLGFFPLVDDTI 304
 Db 226 DGTAVGAPLAHTAIVYKCVSGCSGSDVLAAMDALDGDVILSLWSLGGPIPHRDNI 285
 QY 305 AIGTFRAMERGISYCAAGNGPIESSVANTAPWSTIGAGTLDRFPFAPVRLANGKLLY 364
 Db 286 AIGAYSATERGILVSCSAGNSGSPFITAVNTAPWLTVGASTLDRKIKATVKLNGGEETE 345

QY 365 GESLYPK-----GIKNAGREVEVIYVTGGDKSGSEFCRLRSLPREEIRGKWIC 413
 Db 346 GESAYRPKISNATFTFLDAAKNAKDPSETPY-----CRRGSLTDPAIRGKVLVC 395
 QY 414 DR-QVNGRSRKGAVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVN 472
 Db 396 SALGHVANVKGQAVKDAGGVGMIIINPSQYGVTKSADAHVLPALVVSADGKILAYMN 455
 QY 473 ATVKPKARIIPGGTVIGRSRAPEVAQPSARGPSLANPSILKPDMIAPGVNIIAAMPQNLG 532
 Db 456 STSPVATIAFOGTIIIGDKNAPVAAFPSSRGPSRASGILKPDIIIGGANILAAWPTSD 515
 QY 533 PTGILPYDSRRVNTVMGTSMSCPHVSIGITALIRSYPNWSPAAIKSALMTTADLYDRQ 592
 Db 516 DN-----KNTKSTENIIISGTSMSCPHLSGVAALLKCTHPDWSPAVKSAMMTADTLNAN 571
 QY 593 KAIKDGK-KEAGVPAIGAGHVNPKAINPGLVNIQPDVYITVLTGFTSRDIATHTK 651
 Db 572 SPILDERLLPADIIAIGAGHVNPSRANDPGLVYDTFPEDVLPYLCGLKYTDQVGNLIQR 631
 QY 652 NVSCN---GILRNKPNPFLNYPSTIAVIFKRGKTKTEMITRRVTNVGSPNSIYSVNVKAPEG 708
 Db 632 RVNCSEVKSIIEA---QLNVPSPS-IFGLGSTPTQTYTRTVTVNVGDATSSYKVEVASPEG 686
 QY 709 IKVINPKRIVKHVDOTLSYRVFVVKKNRGKGVASPAQGLTWTY-NSHNLMOVRSP 767
 Db 687 VAIEVERSELNSELNQLTYQVTF---SKRTSSNPEVIEGFLKWTSSNHS-----VRSP 739
 QY 768 ISV 770
 Db 740 IAV 742

RESULT 12
 T06580
 subtilisin-like proteinase (EC 3.4.21.-) p69f - tomato
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
 C:Accession: T06580
 R:Meichtry, J.; Amrhein, N.; Schaller, A.
 submitted to the EMBL Data Library, May 1998
 A:Description: The gene family of subtilisin-like proteases in tomato.
 A:Reference number: Z15771
 A:Accession: T06580
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-747 <MEI>
 A:Cross-references: EMBL:AJ005173; PIDN:CAA06414.1
 A:Experimental source: cultivar VFW8
 C:Genetics:
 A:Gene: p69f
 C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
 C:Keywords: hydrolase; serine proteinase

Query Match 37.7%; Score 1513.5; DB 2; Length 747;
 Best Local Similarity 42.5%; Pred. No. 9e-97;
 Matches 331; Conservative 129; Mismatches 267; Indels 51; Gaps 16;

QY 6 FFLCIIFLLFCCSSSEILQK--QTYIVQLHNPNSETAKTFASKFD---WHLSFLOEAVLG 60
 Db 3 FLKILLIFICSFLRPSIQSDLETVIVHVSPEPNQISTQSSLTDLSEYLSFLPKTTAI 62
 QY 61 EEEBEPSSRLLYSGAIEGFAAQLTESAEILRYSPEVVAVRPHVLQVQTYTSYKFL 120
 Db 63 SSSGDEEAAASMIYSYHNVMKGFAARLTAQVKENEKKGHFVSACKQRIILSHLTHTPSFL 122
 QY 121 GLDGFSGWWSKRFQGTIIIGVLDTGVPSPSPDGTGMPSPRKGICQEGESFSS 180
 Db 123 GLQ--QNGVWKNDSYKGVIGVLDGILPDHPSFDVGMGPPPAKWKVC---ESNFT 177
 QY 181 SSCNKLITGARFFTRGHRVANSPEESPMPREYISARDSTGHGHTHTASTVGGSSVSMANV 240

```

Db 178 TKNNKLIGARSYQLONG-----SPIDNNGHGTHTAGTAAGAFVKGANI 221
Qy 241 LGNGAGVARGMAPAHAIVYKVCWFN-GCVSSDILAAIDVAIQDKVDLSLGLGFPPIPL 299
Db 222 FGNANGTAVGVALAHIAIVYKVCSSDGGSDSDILAMPAAIDDDGVDILSLSGSTKFF 281
Qy 300 YDDTIAIGTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGAGTLDLRRFPVAVRLAN 359
Db 282 HDDGIALGTYSATFGRGIFVSASAGNSGPSLGTVANEAPWILTVGASTHDKLKVTVKLG 341
Qy 360 GKLLYGESLPYKGIKNAGREVEIVYTGDKGSE-----FCLRGLSPREEIRGKVICDR 415
Db 342 SEEFEGESAYHP---KTSNSTFFPLDAGKNESDQFSAFPCSPGSLNDPAIRGKIVLCRL 398
Qy 416 GVN-GRSEKGEAEVKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYVAT 474
Db 399 SISLLRVAQGSYKVDAGGVNMLINEQEGVTYKSAEAFVLPALDVSNADGKILAYMNS 458
Qy 475 VKPKARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAQPONLQPT 534
Db 459 SNPVASTTFHGTGVDGNAPIVASFSSRGPSVASPGLKPDIIIGPVVWVLAAPTSVDNN 518
Qy 535 GLPYDSRRVNTVWGTSMSCPHVSGITALLIRISAYPNWSPAALKSALMTTADLYDRQKA 594
Db 519 ----KNTKSTFNIVSGTSMSCPHLSGVAAALLKGAHPDWSPAALKSAMMTTADTVNLANS 574
Qy 595 IKGNK-PAGVFAIGAGHVPQKAINPGLVYNIQPDVYITLCTLGTRSDILAIHKV 653
Db 575 ILDERLISADLEFAGAGHVPNSASDPLVDYTFPFDYIPYLCGLNTYREVKVLDQKV 634
Qy 654 SCNGILKNPGFSLNYPISIAVIFKRGKTTETMIRRTVNVGSPNSIYSVNVKAPGKIVT 713
Db 635 NCSEV-KRIPEGQLNYPFSI--RLGSTPQTYRTVTVNGDAKSSYKVEIVSPKGVVVKV 691
Qy 714 NPKRLVFKHVDQTLVSRVWFLKKNRGGKVASFAOQLTW-VNSHNLQVRVRSPIV 770
Db 692 EPSALNFTLNQKLYQVIFT---KTNISTTSDVEGFLKWNNSRHS----VRSPIAV 742

RESULT 13
JC6119
subtilisin-like proteinase (EC 3.4.21.-) - tomato
N:Alternate names: calcium-activated endopeptidase; proteinase p69
C:Species: Lycopersicon esculentum (tomato)
C:Date: 23-Mar-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
C:Accession: JC6119
R:Tornerio, P.; Conejero, V.; Vera, P.
Proc. Natl. Acad. Sci. U.S.A. 93, 6332-6337, 1996
A:Title: Primary structure and expression of a pathogen-induced protease (PR-P69) in tom
A:Reference number: JC6119; MUID:96270538; PMID:8692815
A:Accession: JC6119
A:Molecule type: mRNA
A:Residues: 1-745 <TOR>
A:Cross-references: EMBL:X95270; NID:q1524114; PIDN:CAA64566.1; PID:G1524115
C:Comment: This enzyme, a secreted calcium-activated endopeptidase, is a plant pathogene
fense against attacking pathogens.
C:Genetics:
A:Gene: p69
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:137-546/Domain: subtilisin homology #status atypical <SBT>

Query Match 37.1%; Score 1490; DB 2; Length 745;
Best Local Similarity 42.5%; Pred. No. 3.8e-95;
Matches 337; Conservative 111; Mismatches 267; Indels 78; Gaps 19;

Qy 6 FFLCIIILLPCSSSEILQK--QTYIVQLHPNSETAKTPASKPD---WHLSFLQEAVLGV 60
Db 3 FLKILLVIFCSFPWPTIQSNLETYLVHVSPELSLTSQSLDLDLSYLSFLKPTTAT 62
Qy 61 EEEBEPSSRLYXSVGAIEFAQLTSEAEILRYSPEVVAVRPHVLOVQTYGVKFL 120
Db 63 SSSGNEEAATMIYSHNVMTGFAARLTAEQVKEKIHGFVSAQKQRTSLDITHTSSFL 122

```

```

Qy 121 GLDGFNGSGWMSKSRFGQGTIIIGVLTGVWPBPSPDDTGMPISIPRKWKIGICQEGSFSS 180
Db 123 GLQ--QNGWGVKDSNYKGVIIIGVITDGLPDPSPSDVGMPPPPAKWKGVC---ESNFT 177
Qy 181 SSCNKLIGARFTIRGHRVANSPEESPMPREVIISARDSTGHTGHTASTVGGSSVSMANV 240
Db 178 NCKNNKLIGARSYQLGHH-----SPIDDDGHGHTASTAAGAFVNGANV 221
Qy 241 LGNGAGVARGMAPAHAIVYKVCWFNGCYSSDILAAIDVAIQDKVDLSL-LGGFPPIPL 299
Db 222 FGNANGTAAAGVAPPAHAIVYKVCNSDGCADTVLAAMDADIDGVDILSLSLGGGSSDF 281
Qy 300 YDDTIAIGTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGAGTLDLRRFPVAVRLAN 359
Db 282 YSNPIALGAYSATFGRGILVSCSAGNNGPSTGSGVNEAPWILTVGASTQDRKLVKATV 341
Qy 360 GKLLYGESLPYKGIKNAGREVEIVYTGDKGSEFCLRGLSPREEIRG 408
Db 342 REEFEGESAYHPKTSNSTFFALFDAGKNASDEFETPY-----CRSGSUTDPVIRG 391
Qy 409 KMWICDSGVN-GRSEKGEAEVKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLL 467
Db 392 KIVICLAGGVPRVDKQAVKDAGGVGMIIINQORSVTKSAHAHVIPALDISDADGTKI 451
Qy 468 KAYVNAVTKPKARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIA 527
Db 452 LAYVNSTSNVATITFGGTIIGDKNAPIVAASFSSRGPSGASIGILKPDIIIGPVNIIAA 511
Qy 528 PONLPGTGLPYDSRRVNTVWGTSMSCPHVSGITALLIRISAYPNWSPAALKSALMTTAD 587
Db 512 PTSVDNN---KNTKSTFNIIISGTSMSCPHLSGVAAALLKGAHPDWSPAALKSAMMTTAD 567
Qy 588 YDROGKAIKGN-KPAGVFAIGAGHVPQKAINPGLVYNIQPDVYITLCTLGTRSDIL 646
Db 568 LNLANSPLDLRLPADIYALGAGHVPNSRANDEPLVDYTFPFDYIPYLCGLNTYREVG 627
Qy 647 AITHKNVSCN---GILRKNPGFSLNYPISIAVIFKRGKTTETMIRRTVNVGSPNSIYSV 703
Db 628 NLLQKNCSEVKSILEA---QLNYPFS-IVDLGSTPQTYRTVTVNGDAKSSYKVEV 682
Qy 704 KAPEGKIVVNPKLIFK---HVDQTLVSRVWFLKKNRGGKVASFAOQLTW-VNSHN 759
Db 683 ASPEAL-----PSKLTLRANPSSDKLTYQVTF---SKTANSSNTEVIEGFLKWTNRHS 734
Qy 760 LMQRVRSPIVTL 772
Db 735 ----VRSPIALLL 743

RESULT 14
TS1335
subtilisin-like proteinase AIR3, auxin-induced [imported] - Arabidopsis thaliana (fragm
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 03-Nov-2000
C:Accession: TS1335
R:Neuteboom, L.W.; Ng, J.M.Y.; Kuyper, M.; Clijdesdale, O.R.; Hooykaas, P.J.J.; van der
Plant Mol. Biol. 39, 273-287, 1999
A:Title: Isolation and characterization of cDNA clones corresponding with mRNAs that ac
A:Reference number: Z25377; MUID:99178779; PMID:10080694
A:Accession: TS1335
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-758 <NEU>
A:Cross-references: EMBL:AF055848; PIDN:AAC62611.1
A:Experimental source: root culture
A>Note: accumulates during auxin-induced lateral root formation
C:Genetics:
A:Gene: AIR3
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology

Query Match 36.0%; Score 1445.5; DB 2; Length 758;
Best Local Similarity 41.3%; Pred. No. 4.8e-92;

```

Matches	326;	Conservative	120;	Mismatches	274;	Indels	69;	Gaps	19;
Qy	18	SSSEIL---	QKQTYIVQLHPNSETAKTFASKFD----	WHLFSFLOEAVLQVEEEBEPSSR	70				
Db	5	SSKHILASKDSS	YVYFGAHSYGEITEDAMRVKETHYDFLG----	SFTGSRERATDA	60				
Qy	71	LLYSYCSALEGFAAOLTESEAEILRYSPEVAVRPHVLQVQTYTSYKFLGDG----	FGN	127					
Db	61	IFYSTKINGFAAHLHDHLAYESKPEVVSVPFNKALKLHTTRSDWDFLGHNSYYPS	120						
Qy	128	SGVWSKSRFGQGTIIIGVLDTGWPSPESFDDTQWPSITPRKWKIGICQEGSFSSSSCNKL	187						
Db	121	SIWRKARFGEDTIIANLDTGWPSPESKFRDEGLGPIPSRWKIGICQ-NQKDATFHCNKL	179						
Qy	188	IGARPFIRGRVA-----NSPSESNMPREVISARDSTGCHTHTASTVGSSSVSMANVLGN	243						
Db	180	IGARYFNKGYYAAAVGHLNSSFDP-----RDLGHSHTLSAAGDFVPGVYSIFGQ	230						
Qy	244	GAGVARGMAPGAHIAVYKCVW----FNGCYSSDILAAIDVAIOKVDVLSLSLGQFFPL	299						
Db	231	GNGTAKGSPRARVAAYKVCWPPVKGNCEVDADVLAAPDAAIHGADVISVSLGGEPTSP	290						
Qy	300	YDDTIAIGTFRAMERGISVICAGNGNPIESSVANTAPWVSTTIGACTLDRRFPVAVRLAN	359						
Db	291	FNDSSVAIGSFHAACKRIIIVWCAGNSGPADSTVSNVAPQITVGASTMDREFASNVLGN	350						
Qy	360	GKLLVGESL-----YPKGKIKNAGREVEVIYTGDKGSEBCELSGSLPREIRGK	409						
Db	351	GRHYKGQSLSSPALPHAKFYPMASVNAKAK-----NASALDAQCKLGLSLDPIKTKGK	404						
Qy	410	MVICDRGVNGREKGEAKVEAGGVAMILANTEINOEDSDIVHLLPATLIGYTESVLLKA	469						
Db	405	ILVCLRGQNGRVEKGEAVLGGIGCWLENTYVTGNDLLADPHVLPSTQLTSKDSFAVSR	464						
Qy	470	YNATVVKPARIIFGGTVIGSRAPAEVAQPSAKGPSLANPSIIKPDNIIAPGWNIIAANPQ	529						
Db	465	YMTQTKPTAHITPRTDGLKPAQVMASSPSKGPSIVAPQILKPDIITAPGVSVAAYTG	524						
Qy	530	NLGPGLPYDSRRVNTMSGTSMSCPHVSGITALIRSAYPNWSPPAAIKSALMTTADLYD	589						
Db	525	AVSPTNQEPDRRLFLNIAISGTSMSCPHISGIAGLLKTRYPSPAPAIKSAIMTTATTMD	584						
Qy	590	ROGKAIKDG-NKPPAGVFAIGAHGVNPQKAINPGLVNIQPDVIITYLCTLGTRSDILAI	648						
Db	585	DIEPGPIONATNMKATPFSFGAGHVQPNLAVNPGLVYDLGDKYDLNCLSLGYNASQISVF	644						
Qy	649	THKNVSCNGLIKRNGFGS---LNPSTAV-IFRKGKTEMITRRVTVNVCSPNSIIVSNVK	704						
Db	645	SGNNFTCS-----SPKISLVNLNYPSTIVPNLNTSSKYT--VSRVTKVNGRPP-SMTYTKVN	696						
Qy	705	APSGIKVIVNPKRLVFKHVDTLSYRVWFVLKKKNGGKVASFAQGLQTW-VNSHNLQMR	763						
Db	697	NPHGVTVALKPTSLNFTKVGELKTFKVLIV---KSKGNVAKGYMFELVNSAKHY----	749						
Qy	764	VRSPISVTL	772						
Db	750	VRSPIVVKL	758						

RESULT 15
S52769

C22769
 subtilisin-like proteinase ag12 (EC 3.4.21.-) - alder
 C:Species: Alnus glutinosa (alder)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-May-2000
 C:Accession: S52769
 R:Ribeiro, A.; Akkermans, A.D.L.; van Kammen, A.; Bisseling, T.; Pawlowski, K.
 submitted to the EMBL Data Library, March 1995
 A:Description: A nucleotide-specific gene encoding a subtilisin-like protease is involved in
 A:Reference number: S52769
 A:Accession: S52769
 A:Molecule type: mRNA
 A:Residues: 1-761 <RID>
 A:Cross-references: EMBL:X859975; NID:q757521; PIDN:CAA59964.1; PID:q757522

C:Genetics:
A:Gene: agl2
C:Superfamily: subtilisin-like proteinase agl2; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:136-551/Domain: subtilisin homology #status atypical <SBT>

Query March 31.9%; Score 1282; DB 1; Length 761;
Best Local Similarity 38.4%; Pred. No. 1e-80;
Matches 302; Conservative 13; Mismatches 291; Indels 56; Gaps 24;

QY 6 PFLGCIIFLLFCSSEILQOTYIVQJHPNSETAKTFASKDFDHLSPLOQEAFLGVEE 65
DB 15 FASCIICLALHASSPT--MEXSTYIVHM-DKSHMFKAPTSHENMYSS-----IVDCLNSEK 66
QY 66 EPSRLIYSGSALIEGFAAQITSEABILRYSEVVAVRPDHVLQVOTTVSYKFLGIDGF 125
DB 67 PTTSSFYVYTNHVLHGFSASLSHQELDTLRSPGVSAAYDRKATLDTTHTPFLSLNPT 126
QY 126 GNSGVKSWKSRFGOGTIIIGVLDTGVWPPSPDFDGTMPSP-IPRWKKGICQGESFSSSCN 184
DB 127 G--GLWPASNYGEVILIGVDSGVWPPSDSEFKDDGWTQAQVPARKWGLCSR-EGFNSSMCN 183
QY 185 RKLICARFFIRGHRVANSPEESPMPREYISARDSTGCHGTHTASTVGGSSVSMANVLNG 244
DB 184 SKLIGARYFNNGIWA-----IPNATFSMNSARDTLGHTGHTASTAAGNVNGASYEGYG 238
QY 245 AGVARGMAPGHIAYVKVWFNGCYSSDIIAAIDVAIQDVVLVSLGFPFPPLYDDTI 304
DB 239 KGTARGIAPRARVAVYKVTWPEGRYTSDDLQAGIDQAIADGVYDVISISLGDGVPELYEDPI 298
QY 305 AIGTFRAMEGIGSICVCAAGNPIESSVANTAPWSTIGAGTLDRRRPAPVVRVLANGKLY 364
DB 299 AIASFAAMEKGVVSTSGAGNAPFGNWHGIPWLTVAGNIDRSPAGTILTLGNDQTIT 358
QY 365 GESLYPGKGIKNAGREVEIYVTGGDKGSEFLRGLSPREIRKMWICDRGVNGRSEKG 424
DB 359 GWTWFPASATIESS--QLVY---NKTISACNSTELSDAVY-SVVICE-AITPIYAQI 409
QY 425 EAV--KRAQGVAMTLANTEINQESDSDIVHLPLPATLIGYTESVLLKAYVNAIVKPKAII 482
DB 410 DAITRSNVAGAILISNHTKFLFELGGVY---CPLVISPDKAAALIKYAKTDEFFLAGLK 466
QY 483 FGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAAMPQ-----LGPTGLP 537
DB 467 PQEITITGTPAPAVAYYSRRGPSYSGILKPDVWAPGSLVLAWSIPEATAQIG-TNYY 525
QY 538 YDSRRVFTVMSGTSMSCPHVSGITAIRSAYPNWSPAALKSALMTTADLYDROGKAIKD 597
DB 526 LSS---HNMVYSGTSMACPHASGVAALLKAAHPSPAPAIIRSAMMTTANPLDNLNPIHE 582
QY 598 GNKP---AGVFAIGAGHVNPOKAINPGLVYNIQPDVITVYLTCLGFTESDILATHK-NV 653
DB 583 NGKXFHLASPLAMGAGHIDPNRALDPLGLVYDATPDQYINLLCSNMYNKAQIALIVRSDSY 642
QY 654 SCNGILKNPGLFSLNYPISIAVIEKRG---KTETMITRRVTNVGSPNSIYSVNVKAPEGIK 710
DB 643 TCS-----NDPSSDLNYPGF-IAFHNSTCRSRVNTFQRTVNVGGAATYKATVTAPKDSR 697
QY 711 VIVNPKRLVFKHVDQTLISRYVWFLVKKKNRGKGVASFAQGQLTWVNSHNLMMQVRSPISV 770
DB 698 VIVSPQTLAFGSKYKQSYNLTIIINFTRDTKRKQISF--GALVWAN-ENGKHMVRSPIW 754
QY 771 T-LKTN 775
DB 755 SPLRN 760

Search completed: February 2, 2004, 09:11:59
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 09:07:12 ; Search time 17 Seconds
(without alignments)

2143.864 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MEKPFPLCIIIFLLFCSSSS.....NSHNLQVRSPISVTLKTN 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	430	10.7	806	1 SUBV_BACSU	P29141 bacillus su
2	271.5	6.8	382	1 SUBT_BACAM	P00782 bacillus am
3	264.5	6.6	361	1 ELYA_BACHD	P41363 bacillus ha
4	264.5	6.6	1902	1 P2P_IACPA	Q02470 lactobacill
5	261.5	6.5	1181	1 SCAZ_STRPY	P58099 streptococc
6	259	6.4	381	1 SUBT_BACST	P29142 bacillus st
7	257.5	6.4	1167	1 SCAL_STRPY	P15292 streptococc
8	257.5	6.4	1902	1 P3P_IACLC	P15292 lactococcus
9	257	6.4	381	1 SUBT_BACSU	P04189 bacillus su
10	257	6.4	1902	1 P1P_IACLC	P16271 lactococcus
11	255	6.3	381	1 SUBT_BACSA	P00783 bacillus su
12	254	6.3	381	1 SUBN_BACNA	P35835 bacillus su
13	252.5	6.3	1902	1 P2P_IACLC	P15293 lactococcus
14	248	6.2	380	1 ELYA_BACAO	P27693 bacillus al
15	248	6.2	380	1 ELYA_BACCS	P41362 bacillus cl
16	237.5	5.9	274	1 SUBD_BACLI	P00781 bacillus li
17	237.5	5.9	379	1 SUBT_BACLI	P00780 bacillus li
18	229	5.7	513	1 AQLI_THBAQ	P08594 thermus aqu
19	219.5	5.5	269	1 PRTM_BACSP	Q99405 bacillus sp
20	217	5.4	1398	1 PLS_PYRFU	P72186 pyrococcus
21	216.5	5.4	384	1 PRTK_TRIAL	P06873 tritirachiu
22	215.5	5.4	269	1 SUBT_BACLE	P29600 bacillus le
23	213.5	5.3	420	1 SUBT_BACSU	P28842 bacillus sp
24	213.5	5.3	1433	1 SUBF_BACSU	P16397 bacillus su
25	213	5.3	603	1 BPRV_BACNO	P42779 bacteroides
26	211	5.3	275	1 SUBT_BACPU	P07518 bacillus pu
27	211	5.3	319	1 ISPI_BACSU	P11018 bacillus su
28	209.5	5.2	269	1 SUBB_BACLE	P29599 bacillus le
29	206.5	5.1	388	1 CUDP_METAN	P29138 metathizium
30	205.5	5.1	595	1 BPRX_BACNO	P42780 bacteroides
31	204	5.1	378	1 ELYA_BACSP	P20724 bacillus sp
32	200	5.0	387	1 PRTR_TRIAL	P23653 tritirachiu
33	199.5	5.0	580	1 EXPR_XANCP	P23314 xanthomonas

34 199 5.0 1052 1 MS1P_HUMAN Q14703 homo sapien
35 198 4.9 1052 1 MS1P_CRIGR Q22248 cricetus
36 196.5 4.9 467 1 ISP6_SCHPO Q40903 schizosacch
37 196.5 4.9 534 1 PROA_VIBAL P16588 vibrio algi
38 196 4.9 1052 1 MS1P_RAT Q9wt23 rattus norv
39 195 4.9 1052 1 MS1P_MOUSE Q9wt22 mus musculu
40 194 4.8 326 1 ISP_FAEPO P29139 paenibacill
41 193 4.8 422 1 TKSU_PYRKO P58502 pyrococcus
42 184 4.6 408 1 SEPR_THESR P80146 thermus sp.
43 183 4.6 533 1 PEPC_ASPNG P33295 aspergillus
44 181 4.5 409 1 ALP_TRIHA Q03420 trichoderma
45 178 4.4 401 1 THES_BACSP Q45670 bacillus sp

ALIGNMENTS

RESULT 1
SUBV_BACSU
ID SUBV_BACSU STANDARD; PRT; 806 AA.
AC P29141;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor extracellular protease vpr precursor (EC 3.4.21.-).
GN VPR OR IPA-45R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
RX MEDLINE=92041574; PubMed=1938992;
RA Sloma A., Rufe G.A. Jr., Theriault K.A., Dwyer M., Wilson S.W.,
RA Pero J.;
RT "Cloning and characterization of the gene for an additional
RT extracellular serine protease of Bacillus subtilis";
RL J. Bacteriol. 173:6889-6895(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.F.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiesch J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mollado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Mambutt R., Wedler H., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*,"
RL Nature 390:249-256 (1997).
CC -!- FUNCTION: NOT REQUIRED FOR GROWTH OR SPORULATION.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; M76590; AAA22881.1; -;
CC EMBL; X73124; CAA51601.1; -;
CC EMBL; Z99123; CAB15835.1; -;
CC PIR; A41341; A41341;
CC HSP; P00782; ZSBI.
CC MEROPS; S08.09A; -;
CC Subtilisin; BG10591; vpr.
CC InterPro; IPR001137; PA.
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF02225; PA; 1
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE_ASP; 1.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 160
FT CHAIN 161 806 MINOR EXTRACELLULAR PROTEASE VPR.
FT ACT_SITE 189 189 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 233 233 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 534 534 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 806 AA; 85608 MW; F984E3BF0B869DDD CRG64;
Query Match 10.7%; Score 430; DB 1; Length 806;
Best Local Similarity 23.2%; Pred. No. 1-2e-21;
Matches 198; Conservative 135; Mismatches 299; Indels 222; Gaps 35;
QY 7 FLCIFLLPCSSSEIL-----ETAKTFASKFDWHLFLQEAIVLGVVEEESPSRLLSYSGAIG 36
DB 8 FLVSVFLPFLALSTGIVGQVQAPASSKTSADLEAEVFGDIDMTTSKTTIVIVELKESL 67
QY 37 -----ETAKTFASKFDWHLFLQEAIVLGVVEEESPSRLLSYSGAIG 81
DB 68 AEAKEGESQSKLKTARTKAK-----NKAIKAVK-----NGKNREYEQVFGS 112
QY 82 FAQLTSEAEILRYSEVAVRP-----DHVLQVQTTYSYKFLG--LDG-----FGNSG 129
DB 113 FSKMLPANEIPKLLAVKDKVAVPNTVYTKDNNKDKDVTISEDAVSPQMDGAPYIGAND 172
QY 130 VMSKSRFGQGTIIIGVLDITGVWPSPSFDDTCMPISIPRKWKIGIQEGESFSSSSCNRLKIG 189
DB 173 AWDLYTGKIGKVAIDITGVYHNPD-----KKNFG 204
QY 190 ARFFIRGHRVANS-----PEESP-ANPREYISARSTGHTASTVGGSSVSMANVLGNA 245
DB 205 Q---YKGYDFVNDYDPKETPTGDP-----GEATDGHGTAVGVAAN----- 244
QY 246 GVARGMAPGAHIAVYKVCWNGCVSSD-ILAAIDVAIQDKVDVLSLGGFFPIPLDYDTI 304

Db 245 GTIKGVAPDATTLLAYRVLPFGSGGTENTVIAGVERAVQGDVMMNLGN---SLNNPDW 301
QY 305 AIGTFR--AMERGIVICAGNNGPIESSVAN--TAPWVSTTIGAGTLDRRFPVAV--RLA 358
Db 302 ATSTALDWNASEGVAVTSNGNSGPNGTWVGSPGTSREALSVCATQLPLNEVAVTGGSYS 361
QY 359 NGKLLYGESLYPGKIGKAGREVEVIYVVGDKGSEFCRLGSLPREIRGKWKVICDRGVN 418
Db 362 SAKVMGYNKEDDVKALNN--KEVELV-EAGIGSEAKDF-----EGKDLTGKVAVVKRGS 412
QY 419 GRSEKGEAVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAVYNATVKPK 478
Db 413 AFVDKADNAKGAIGAKVWNNLSGSEIANVPGMSVPTIKLSDEGEKL---VSALKAGE 459
QY 479 ARIIFGTVIGRGAPEVAQFSARGPSLANPSILKEDMIAPGVNIIAAMPQNLGPTGLPY 538
Db 470 TKITFKLTV--SKALGQVADFSSRGP-VMDTWMIKPDISAPGVNIVSTIPTDHPHPYGY 527
QY 539 DSRVNTVMGSGTSMGCPHVSIGTALIRGAYPNWSPAAIKSALMTTA-DLYDRQKAIKD 597
Db 528 GSKQ-----GTSMASPHIAGAVAVIKQAKPKWSVEQIKAAIMNTAVTLKDSGDGEVPH 580
QY 598 GNRPAAGVFAIGAGVNPQKAINPLVYNIQPVVDYITVLTGLG-FTRSDILAIITHKN--- 652
Db 581 N-----AQGAGSARIMNAIKADSLVSPGYSYGTFLKENGNETKNETFTTENQSSIRK 633
QY 653 -----VSCNGILRKNGPFSL-----NYPSTAVIFKSGKT- 681
Db 634 SYLSEYFNGSGISTGTSRVIPAHQTKATAKVKVNTKTKAGTYEGTVIVREGKTV 693
QY 682 -----TEMITR-----RVTVN-----GSPNSIYVNVKAPEG-----IKVIVNPKRL- 718
Db 694 AKVPTLLIVKEPDYPRVTSVSVSEGSVQGTYQIETVLPAGAEALFLVYDNLDFAGQAG 753
QY 719 VFKAVDQTLRYRW 732
Db 754 IYKNQDKGYQYFDW 767
RESULT 2
SUBT BACAM
ID SUBT BACAM STANDARD; PRT; 382 AA.
AC P00782;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Subtilisin BPN precursor (EC 3.4.21.62) (Subtilisin Novo) (Alkaline
DE protease).
GN APR.
OS *Bacillus amyloliquefaciens*.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.
OX NCBI_TaxID=1390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23844;
EX MEDLINE=85006739; PubMed=6090391;
RA Vasantha N., Thompson L.D., Rhodes C., Banner C., Nagle J.,
RA Filpula D.;
RT "Genes for alkaline protease and neutral protease from *Bacillus*
RT *amyloliquefaciens* contain a large open reading frame between the
RT regions coding for signal sequence and mature protein.";
RL J. Bacteriol. 159:811-819 (1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84069812; PubMed=6316278;
RA Wells J.A., Ferrari E., Henner D.J., Estell D.A., Chen E.Y.;
RT "Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens*
RT subtilisin in *Bacillus subtilis*.";
RL Nucleic Acids Res. 11:7911-7925 (1983).
RN [3]
RP SEQUENCE OF 108-382.
EX MEDLINE=68086682; PubMed=6065094;
RA Markland F.S., Smith E.L.;

RT "Subtilisin BPN. VII. Isolation of cyanogen bromide peptides and the
 RL complete amino acid sequence.";
 RN J. Biol. Chem. 242:5198-5211(1967).
 [4]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=72035041; PubMed=4399039;
 RA Alden R.A., Wright C.S., Kraut J.;
 RT "A hydrogen-bond network at the active site of subtilisin BPN.";
 RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 257:119-124(1970).
 [5]
 RN X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF COMPLEX WITH INHIBITOR.
 RP MEDLINE=85033707; PubMed=6387152;
 RX MEDLINE=90057412; PubMed=2684274;
 RA Pantoliano M.W., Whitlow M., Wood J.F., Dodd S.W., Hardman K.D.,
 RA Röllence M.L., Bryan P.N.;
 RT "Large increases in general stability for subtilisin BPN' through
 RL incremental changes in the free energy of unfolding.";
 RN Biochemistry 28:7205-7213(1989).
 [6]
 RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF MUTANT.
 RP MEDLINE=90057412; PubMed=2684274;
 RX Pantoliano M.W., Whitlow M., Wood J.F., Dodd S.W., Hardman K.D.,
 RA Röllence M.L., Bryan P.N.;
 RT "Large increases in general stability for subtilisin BPN' through
 RL incremental changes in the free energy of unfolding.";
 RN Biochemistry 28:7205-7213(1989).
 [7]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX Gallagher T., Oliver J., Bott R., Betzel C., Gilliland G.L.;
 RT "Subtilisin BPN' at 1.6-A resolution: analysis for discrete disorder
 RL and comparison of crystal forms.";
 RN Acta Crystallogr. D 52:1125-1135(1996).
 [8]
 RP ACTIVE SITE.
 RX MEDLINE=69104413; PubMed=5249818;
 RA Markland F.S., Shaw E., Smith E.L.;
 RT "Identification of histidine 64 in the active site of subtilisin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 61:1440-1447(1968).
 CC -I- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
 CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 CC for peptide bonds, and a preference for a large uncharged residue
 CC in PL. Hydrolyzes peptide amides.
 CC -I- SUBCELLULAR LOCATION: Secreted.
 CC -I- BIOTECHNOLOGY: Used as a detergent protease. Sold under the name
 CC Alcalase by Novozymes.
 CC -I- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
 CC SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
 CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
 CC IS NOT NECESSARY FOR NORMAL SPOULATION.
 CC -I- SIMILARITY: Belongs to peptidase family S8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; K02496; AAB05345.1; -.
 DR PIR; B25415; SUBSN.
 DR PDB; 1S01; 15-OCT-90.
 DR PDB; 1S02; 15-JAN-92.
 DR PDB; 1SBH; 07-DEC-95.
 DR PDB; 1SBI; 07-DEC-95.
 DR PDB; 1SEN; 31-JAN-94.
 DR PDB; 1SBT; 31-MAY-84.
 DR PDB; 2SBT; 31-MAY-84.
 DR PDB; 1SIB; 31-OCT-93.
 DR PDB; 2SIC; 15-APR-93.
 DR PDB; 3SIC; 31-JAN-94.
 DR PDB; 5SIC; 31-JAN-94.
 DR PDB; 2SNT; 15-JAN-93.
 DR PDB; 1SPB; 15-OCT-95.

DR PDB; 2ST1; 15-JUL-91.
 DR PDB; 1ST2; 15-JUL-91.
 DR PDB; 1SUA; 14-JAN-98.
 DR PDB; 1SUB; 31-JAN-94.
 DR PDB; 1SUC; 31-JAN-94.
 DR PDB; 1SUD; 31-JAN-94.
 DR PDB; 1SUE; 14-OCT-95.
 DR PDB; 1SUP; 14-NOV-95.
 DR PDB; 1AK9; 12-NOV-97.
 DR PDB; 1AU9; 31-DEC-97.
 DR PDB; 1AQN; 14-JAN-98.
 DR PDB; 1YJA; 11-JUL-96.
 DR PDB; 1YJB; 11-JUL-96.
 DR PDB; 1YJC; 11-JUL-96.
 DR PDB; 1A2Q; 29-APR-98.
 DR PDB; 1DUI; 28-JAN-00.
 DR PDB; 1GNV; 24-JUN-02.
 DR PDB; 1LW6; 21-AUG-02.
 DR PDB; 1UBN; 28-JUN-99.
 DR MEROPS; S08.034; -.
 DR InterPro; IPR000209; Peptidase S8.
 DR Pfam; PF00082; Peptidase S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal;
 KW 3D-structure.
 FT SIGNAL 1 32
 FT PROPEP 33 107
 FT CHAIN 108 382
 FT ACT_SITE 139 139
 FT ACT_SITE 171 171
 FT ACT_SITE 328 328
 FT HELIX 113 117
 FT TURN 118 119
 FT TURN 127 127
 FT TURN 120 126
 FT TURN 131 132
 FT TURN 134 139
 FT TURN 144 145
 FT TURN 147 148
 FT STRAND 151 156
 FT TURN 164 165
 FT HELIX 171 180
 FT TURN 193 194
 FT STRAND 196 201
 FT STRAND 203 203
 FT TURN 205 206
 FT STRAND 209 209
 FT HELIX 211 223
 FT TURN 224 225
 FT STRAND 228 231
 FT STRAND 235 235
 FT HELIX 240 251
 FT TURN 252 253
 FT STRAND 255 259
 FT STRAND 266 266
 FT TURN 267 268
 FT STRAND 269 269
 FT STRAND 274 274
 FT TURN 275 277
 FT TURN 279 280
 FT STRAND 282 287
 FT TURN 289 290
 FT STRAND 293 293
 FT TURN 295 296
 FT TURN 301 302
 FT STRAND 305 308
 FT STRAND 312 316
 FT TURN 317 319
 FT STRAND 320 324
 FT HELIX 327 344

```

FT TURN 346 347
PT HELIX 350 359
PT TURN 360 360
PT STRAND 362 362
FT HELIX 367 370
PT TURN 371 372
FT STRAND 374 374
PT HELIX 377 380
PT TURN 381 382
SQ SEQUENCE 382 AA; 39181 MW; ED987DAFA37B8335 CRC64;

Query Match 6.8%; Score 271.5; DB 1; Length 382;
Best Local Similarity 20.5%; Pred. No. 2.9e-11;
Matches 126; Conservative 63; Mismatches 177; Indels 249; Gaps 17;

QY 6 FFLCIIFLFCSSSEILQKQYIVQLHPNSTATKAFKPDWHLSTFQBAVLGVBEDEE 65
DB 12 FALALIFTMAFGSTSS-----AQAGKSNRKKYIVGFQKTWTSTSAKKKDVISEK 63

QY 66 EBSRLLYGSAIGFAQALTESEAEILRYPEVAVRDPDHLVQVQTTYSYKF-LGLDGG 124
DB 64 GKGKQKQFYK---VDAASATLNEKAVKELKDPSPVAYVEEDVAH---AYAQSVPYGVSQ 117

QY 125 FGNISGWSKSRFGQGTIIIGVLDTGVMPSPSPFDDTCMPSPKRWKGIQCEGESFSSSCN 194
DB 118 IKAPALHSQGYGTSNVKVAIVDSGIDSSHPDLKVAGGASM----- 157

QY 185 RKLIGARFTIRHRVANSPEESPNNPREYISARDSTGHGTHFTASTVGGSSVSMANVLNG 244
DB 158 -----VPSETNPFQDNNSHGTHVAGTVA-----ALNNS 185

QY 245 AGVARGMAPGAHATYVKVWFGNC-YSSDILAAIDVAIQDKVDVLSLGGPFPLYDDT 303
DB 186 IGVLTGVAPASLYAVKVLGADGGQGYWIINGIEWAIANNMDVINMSLGG---PSGSAA 241

QY 304 IAITFRAMERGIVTCAAGNNGPIESSVANTAPVSTIGACTLDRRPPAVVRLANGKLL 363
DB 242 LKAANDKVASGVVVVAAGNETSGSS-----STVG----- 273

QY 364 YGBSLYPGKGIKNAGREVEYIVYTGDKGSEFCLGRSLPREIRGMVICDRGVNGRSEK 423
DB 274 -----YFGK----- 277

QY 424 GEAVKGAGGVAMTLANTEINQBEDSDVHLHPATL-IGYTESVLLKAYNAVTKPKARII 482
DB 278 -----YPSVIAVGAVDS----- 289

QY 483 FGGTVIGRSPAEVAQFSARGPSLANPSILKPDMIAPGVNIIAAMPQNLGPTGLPVDSE 542
DB 290 -----SNQRAGFSVSGPEL-----DVMAPGVSIQSTLPGN----- 319

QY 543 VNFVMSGTSWSPHYSGITALIRSAVPNWSAAIKSALMTTADLYDRQKAIKDNKPA 602
DB 320 -KYGAYNGTSMASPHVAGAAALILSKHPNWTNTQVRSSLENTT-----TKLG 365

QY 603 GVFAIGAGHVNPKA 617
DB 366 DSFYVKGKGLINQAA 380

RESULT 3
ID ELYA BACHD STANDARD; PRT; 361 AA.
AC P41363; 053294; 09KEJ7;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
GN Thermostable alkaline protease precursor (EC 3.4.21.-).
DN BH0855.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=AH-101 / JCM 9161;
RX MEDLINE=93098926; PubMed=1369007;
RA Takami H., Kobayashi T., Aono R., Horikoshi K.;
RT "Molecular cloning, nucleotide sequence and expression of the
RL structural gene for a thermostable alkaline protease from Bacillus
RN sp. no. AH-101.";
RP Appl. Microbiol. Biotechnol. 38:101-108(1992).
RA [2]
RP REVISIONS.
RA Takami H., Kobayashi T., Aono R., Horikoshi K.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RL Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RN Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RL halodurans and genomic sequence comparison with Bacillus subtilis.";
RN Nucleic Acids Res. 28:4317-4331(2000).
RA [4]
RP SEQUENCE OF 94-113.
RC STRAIN=AH-101 / JCM 9161;
RX MEDLINE=91103969; PubMed=1370008;
RA Takami H., Akiba T., Horikoshi K.;
RT "Characterization of an alkaline protease from Bacillus sp. no.
RL AH-101.";
RN Appl. Microbiol. Biotechnol. 33:519-523(1990).
CC -1- FUNCTION: SHOWS KERATINOLYTIC ACTIVITY AND HAS A HIGH OPTIMUM PH
CC (PH 12-13).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D13158; BAA02443.2; -.
CC EMBL; AP001510; BAB04574.1; -.
CC F1; G83756; G83756.
CC HSP; C99405; 1MPT.
CC MEROPS; S08.UPA; -.
CC InterPro; IPR00209; Peptidase S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE ASP; 1.
CC PROSITE; PS00137; SUBTILASE HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
KW SIGNAL 1 24
FT PROPEP 25 93
FT CHAIN 94 361
FT ACT_SITE 124 124
FT ACT_SITE 154 154
FT ACT_SITE 307 307
FT VARIANT 38 38
FT VARIANT 69 70
FT VARIANT 74 77
FT VARIANT 86 86
FT VARIANT 104 104
FT VARIANT 292 292
FT VARIANT 346 347
SQ SEQUENCE 361 AA; 38116 MW; 24BF004F9E3E8474 CRC64;

Query Match 6.8%; Score 264.5; DB 1; Length 361;
Best Local Similarity 19.6%; Pred. No. 8.2e-11;
Matches 115; Conservative 73; Mismatches 141; Indels 259; Gaps 18;

```



```

QY 295 FP--IPLYDDTTAIGTFRAMERGISVICAAGNNGPIESS-----VAN 334
DB 353 DSGNQTLDEPETA-AVQNA NESGTAIVISAGNSGTSATQGVNDYVGLQDNEMVGP 411
QY 335 TAPWVSTIGAGTLDRFPFAVRLANGKLLYGSFLYPGKGNAGREVEIYVYTGDKGSE 394
DB 412 TSGGATVSAENTDVISQAVITDKDL--QLGP-----ETIQISSLNDFGTS 457
QY 395 FCUR-----GSLPR-----BEIRGKVICDRGVNGRSEKGEAVKAGGVAMILA 438
DB 458 FDKQFVYVVKDASGLSGAADAADYADAKGIAIVKRGELNFAADKQVQAAGAGLIIV 517
QY 439 NTHINEEDSIDVHLLPATLIGTVESV-----LLKAYVNAVTKPKARIIFGTV 487
DB 518 NN-----DGTATPLSIRLITTPFFGLSSGTQKLVNDVTAHDDSLGVKIALTL 568
QY 488 IGRSRAPE--VAQFSARGSLANPSILKPDMLAPGVNIIAANFQNLGTLGYDSRRVNF 545
DB 569 LPNQKTEDKMSDFTSYGP-VSNLS-FKPDITAPGNINWSTQNN-----GY 613
QY 546 TVMSGTSMCPHVSIGTALIRISAYPNWSPAAIKSALMTTADLYDRQKAIDGNKPAVGF 605
DB 614 TMSGTSMASPPFTAGSQALLQALANKNPFY-----ADYKQLGTALTDFLKTV-- 664
QY 606 AIGAGHVPQKAINPGLVNIQPDYITLCTLGTRSDILAIHKNVSCNGILKPNFGF 665
DB 665 -----MNTAQPINDINNNVIVSPRGAGLVQVKAIDA-LEKNPST 706
QY 666 SL---NYPISAV-----IFK---RGKTTEMITRRVTVNGSPNSIYVNVKADEGI 709
DB 707 VVAENGYPAVELEKDFSTDKTKFLFTNTHLYQWDSNTDTHAVTISATDPNSGV 764

RESULT 5
SCA2_STRPY STANDARD; PRT; 1181 AA.
ID SCA2_STRPY
AC P38099;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CSA peptidase precursor (EC 3.4.21.-) (SCP).
GN SCFA OR SPY2010.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Szate S., Suvarov A.N., Keston S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus
RT pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -|- FUNCTION. THIS VIRULENCE FACTOR OF S. PYOGENES SPECIFICALLY CLEAVES
CC THE HUMAN SERUM CHEMOTAXIN CSA AT LYS(68)-ASP(69) BOND NEAR ITS
CC C-TERMINUS, DESTROYING ITS ABILITY TO SERVE AS A CHEMOATTRACTANT.
CC -|- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (potential).
CC -|- SIMILARITY: Belongs to peptidase family S8.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL; AE006623; AAK34691.1; -.
DR HSP; Q45670; IDBI.
DR MEROPS; S08.020; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LpxTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; FALSE NEG.
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Repeat;
KW Signal; Complete proteome.
FT SIGNAL 1 31 BY SIMILARITY.
FT CHAIN 32 1147 CSA PEPTIDASE.
FT PROPEP 1148 1181 REMOVED BY SORTASE (POTENTIAL).
FT ACT_SITE 130 130 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 193 193 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 512 512 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1029 1121 CELL WALL ATTACHMENT DOMAIN (POTENTIAL).
FT DOMAIN 1034 1118 5 X 17 AA TANDEM REPEATS.
FT REPEAT 1034 1050 1.
FT REPEAT 1051 1067 2.
FT REPEAT 1068 1084 3.
FT REPEAT 1085 1101 4.
FT REPEAT 1102 1118 5.
FT SITE 1144 1148 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1147 1147 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1181 AA; 129486 MW; 39FC51763419CFC CRC64;

Query Match 6.5%; Score 261.5; DB 1; Length 1181;
Best Local Similarity 24.2%; Pred. No. 6; Se-10;
Matches 128; Conservative 95; Mismatches 198; Indels 107; Gaps 26;

QY 133 KSRFGQGTIIGVLDGTGWPESPESFDDTGMP-SIPRKWKGICQGESFSSSSCNKRLIGAR 191
DB 117 KAGKGAGTVVAVIDAG-----FDKNHEAWRLTDTKARYQSKEDLEKA---KKEHGIT 166
QY 192 FFIGRHVANSPEESPMPREVISARDSTGCHGHTASTVGGSSVSMANVLGNGAGVARGM 251
DB 167 Y---GEWMDKVAHYHDYSKDGKTAVDQE-HGHVSGILSGNAPSETKBPYR-----LEGA 218
QY 252 APGAHIAVYKVCWFNGC---YSSDILAAIDVAIQKVDVLSLSLGGFPPIP---LYDDTIAI 306
DB 219 MPEAQILLMEVINGLADYARNYAQAIDAVNLGAKVINMSFGNAALAYANLPDETKKA 278
QY 307 GTFRAMERGISVICAAGNNG-----PI-----ESSVANT---APWVSTIGAGTLDRFP 351
DB 279 FDY-AXSKGVSIIVTSAGNDSFGGKTRLPDLADHPDYGWVGTPAABDSTLTVASYSPPKQL 337
QY 352 --PAVVLANK-----LLYGESLYPGKGNAGREVEIYVYTGDKGSEFCRLGSLPRE 404
DB 338 TETATVKTADQODKEMPFLSTNFEPNKAYDYA-----YANRGMKEDDF-----K 382
QY 405 EIRGKVICDRGVNGRSEKGEAVKAGGVAMILANTEINQEE-----DSIDVHLLPATLIG 460
DB 383 DVKGKIALIERGDIIDFKDKIANAKKAGAVGLIYD---NODKGFPIELPNVDQMPAFIS 439
QY 461 YTESVLLIKAYVNAVTKPKARIIFGGT--VIGRSRAPEVAQFSARGPSLANPSILKPDMLA 518
DB 440 RKDGLLKE-----NPQKITITENATPKVLPTASGKLSRFSSWG--LTADGNKIPDIAA 491
QY 519 PGVNIIAANFQNLGPTGLPYDSRRVNFVMSGTSMCPHVSIGTALIR-----SAYPNWSP 574
DB 492 PGQDILSSVANN-----KYAKLSGTSMSGAPLVAGIMGLLQKQYETQYDMP 538
QY 575 A---AIKSALMTAD-LYDRQKKAIDGNKPAVFAIGAGHVPQKA 617
DB 539 SERLDLAKKVLMSATALYDEDEKAFSPRQ-----GAGAVDAKKA 580

```

```

RESULT 6
SUBT_BACST
ID SUBT_BACST STANDARD; PRT; 381 AA.
AC R29142;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Subtilisin J precursor (EC 3.4.21.62).
GN APRJ
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 10278 / KCTC 1823;
RX MEDLINE=92231938; PubMed=1567435;
RA Jang J.S., Kang D.O., Chun M.J., Byun S.M.;
RT "Molecular cloning of a subtilisin J gene from Bacillus
RL Biochem Biophys Res Commun. 184:277-282(1992).
CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
CC for peptide bonds, and a preference for a large uncharged residue
CC in pl. Hydrolyzes peptide amides.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
CC SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
CC IS NOT NECESSARY FOR NORMAL SPOULATION.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M64743; AAA22247.1; -.
CC PIR; J01487; J01487.
CC HSP; P04189; ISCU.
CC MEROPS; S08.035; -.
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE_ASP; 1.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC Hydrolase; Spoolation; Serine protease; Zymogen; Signal.
CC SIGNAL 1 29 POTENTIAL.
CC PROPEP 30 106 POTENTIAL.
CC CHAIN 107 381 SUBTILISIN J.
CC ACT_SITE 138 138 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 170 170 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 327 327 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC SEQUENCE 381 AA; 39495 MW; C7A596F7629087D5 CRC64;
Query Match 6.4%; Score 259; DB 1; Length 381;
Best Local Similarity 21.2%; Pred. No. 2.1e-10;
Matches 130; Conservative 61; Mismatches 175; Indels 248; Gaps 20;
QY 6 FFLCIPLFLPCSSSEILQQTIVIVOLHPNSETAKTFASKEDWHLSFLQEAFLVGEVEE 65
DB 12 FAULTIPLTFMFSNWS-----VQAAGKSTTEKKYIVGFKQTMSAMSAKKVDVSK 62
QY 66 EPSRLLYSYGSAIEGFAALTESEAEILRYSPEVAVRPDHLVQLQVTTYSYKFLGLDGF 125
DB 63 GGRVQKQPKY---VNAATLDEKAVKELKQDSVAYVEEDHIAH-EYAQSVPY-GISQI 117

```

```

QY 126 GNSGVWSKSRFGQGTIIIGVLDTCVWPESPFDFTGMPSIPRKWKIGICQEGESFSSSCNR 185
DB 118 KAPALHSQGTGNSVNVAVVIDSGIDSSHPLDNRVG-----
QY 186 KLICAREFFIRGHRVANSPEESPNNPREYISARSTGCTGHTASTVGGSSVSMANVLNGA 245
DB 153 ---GASP-----VPSETNPYQGSSEHGHVAGTIA-----ALNNSI 185
QY 246 GVARGMAPGAHIAVYKVCWENG-YSDDLAAIDVAIQKVDVLSLSLGGFFPLDYDDTI 304
DB 186 GVL-GVSPSASLVAVKVLDTSGSQYSWIINGIEWALSNNMVDVNNLSLG---PSGSTAL 241
QY 305 AIGTFRAMEGIVSICAGNNGPIESVANTAPWSTIGAGTLDLRPPAVRLANGKLLY 364
DB 242 KTVVDKAVSSGIVVAAAAGNEGSGS-----STVG-----
QY 365 GESLYPGKGIKNAGREVEVIYVGGDKSGFCLRGSLPREIRGMVCDRGVNGRSEK 424
DB 273 ---YPAK-----
QY 425 EAVKEAGGVAMILANTEINQEDSIDVHLLPATL-IGYTESVLLKAYVNAVTKPKARIIF 483
DB 277 -----YPSIIAVGAVNS-----
QY 484 GGTVIGSRAPVAQFSARGFSLANPSILKPDMAFGVNIIAAMPQNLGPTGLPYDSRV 543
DB 289 -----SNQASFSAGSEL-----DWAPGVSI-----QSTLPGG-----
QY 544 NPTVMSGTSCSHVSGITALIRSAIPNWSPPAAKSAALMTADLYDRQGAIKDGNKPG 603
DB 319 TYGAYNGTSNATPHVAGAAALILSKHPTWTNAQVRDLSEATYL-----GNS--- 366
QY 504 VFAIGAGHVNPQKA 617
DB 367 -FYVKGGLINQAA 379
SCAL_STRPY STANDARD; PRT; 1167 AA.
AC P15926;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CSA peptidase precursor (EC 3.4.21.-) (SCP).
GN SCPA.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-41.
RX MEDLINE=90153964; PubMed=2406246;
RA Chen C.C., Cleary P.P.;
RT "Complete nucleotide sequence of the streptococcal C5a peptidase gene
RL J. Biol. Chem. 265:3161-3167 (1990).
CC -!- FUNCTION: THIS VIRULENCE FACTOR OF S. PYOGENES SPECIFICALLY CLEAVES
CC THE HUMAN SERUM CHEMOTAXIN C5A AT LYS(68)-ASP(69) BOND NEAR ITS
CC C-TERMINUS, DESTROYING ITS ABILITY TO SERVE AS A CHEMOATTRACTANT.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (potential).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J05229; AAA26960.1; -.

```

```

DR PIR; A35066; A35066.
DR HSP; Q45670; 1DBI.
DR MEROPS; S08.020; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006132; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Repeat;
KW Signal.
FT SIGNAL 1 31
FT CHAIN 32 1130 CSA PEPTIDASE.
FT PROPEP 1131 1167 REMOVED BY SORTASE (POTENTIAL).
FT ACT_SITE 130 130 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 193 193 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 512 512 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1029 1104 CELL WALL ATTACHMENT DOMAIN (POTENTIAL).
FT DOMAIN 1034 1101 4 X 17 AA TANDEM REPEATS.
FT REPEAT 1034 1050 1.
FT REPEAT 1051 1067 2.
FT REPEAT 1068 1084 3.
FT REPEAT 1085 1101 4.
FT SITE 1127 1131 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1130 1130 AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT SEQUENCE 1167 AA; 128263 MW; D2DDC5E5752DA5D CRC64;

Query Match 6.4%; Score 257.5; DB 1; Length 1167;
Best Local Similarity 24.1%; Pred. No. 1.2e-09;
Matches 127; Conservative 96; Mismatches 198; Indels 107; Gaps 26;

QY 133 KSRFGQGTIGVLDTGWPSPSPFDTGWP-SIPKWKICQEGESFSSSCNRKLGAR 191
DB 117 KAGKAGTVAVIDAG-----FDKNHEARLTDKTKARYSKDELEK---KKEHGTT 166
QY 192 FPIRGHRVANSPEESPNMREYTSARDSTGHGHTASTVGGSSVANVLGNAGVARGM 251
DB 167 Y---GEWNDKVAYHYDYSKGTAVDOE-HGTEVSGILSGNAPSTKEPYR-----LEGA 218
QY 252 APGAHIAVYKVFENGCG--YSSDILAAIDVAIQDYVLSLGLGPPFP---LYDDTIAI 306
DB 219 MPEAQLLMREVEIVNGLADYARNVAQAIRDAVNLGAKVINMFGNAALAYANLPDETCKA 278
QY 307 GTFRAMERGISVCAAGNG-----PI-----ESSVANT---APWVSTIGAGTLDRFP 351
DB 279 FDY-AKSKGVSIIVTSAGNDSFSGKTRPLADHPDYGVVGTTPAADSTLTVASYPDKQL 337
QY 352 --PAVVELANGK-----LLYGESLPGKGIKNAGREVEIYVYTGDKGSEFCRLGSLPRE 404
DB 338 TETAMVTDQDQKEMPEVLTNRPENKAYDYA-----YANRGMKEDDF-----K 382
QY 405 EIRGKMWICDRGVNGRSEKGEAVKEAGGVAMILANTEINQEE---DSIDVHLLPATLIG 460
DB 383 DVKQKIALIERGIDDFDKVANAKKAGAGVLYD---NQDKGFPIELFNVQMPAAFTS 439
QY 461 YTESVLLKAYNATVKPARIIFGGT--VIGRSRAPEVAQPSARGPSLANPSILKPDMTA 518
DB 440 RKDGLLLKD-----NPQKTIITFNATPKVLPTASGTKLSRFSSWG--LTADGNKPKDIAA 491
QY 519 PGVNIIAAMPQNLGPTGLPYDSRVNFTVMSGTSMSCPHVSIGTALIR-----SAYPNWSP 574
DB 492 PGQDILSSVANN-----KYAKLSGTSMSAPLVAGIMGLLQKQYETQYPDWTF 538
QY 575 A---ALTKSALMTTAD--LYDRQGGKAIKDKGNKPGAVFAIGAGHVNPOKA 617
DB 539 SERLDLAKVLMSSATALYDEDEKAYFSPRQQ-----GAGAVDAKKA 580

```

```

RESULT 8
P3P_LALCLC
ID P3P_LALCLC STANDARD; PRT; 1902 AA.
AC P15292; 1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PIII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
associated serine proteinase).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN=SK11;
RX MEDLINE=89340435; PubMed=2760036;
RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
cell envelope-located serine proteinase.";
RL J. Biol. Chem. 264:113579-113585(1989).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
specificity, although some substrate preference have been noted,
e.g. large hydrophobic residues in the P1 and P4 positions, and
Pro in the P2 position. Best known for its action on caseins, and
although it has been shown to hydrolyze hemoglobin and oxidized
insulin B-chain.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
an amide bond (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL; J04962; AAA03533.1; ALT_SEQ.
HSPF; P00782; 2SET.
MEROPS; S08.019; -.
InterPro; IPR001899; Gram_pos_anchor.
InterPro; IPR006132; LPXTG.
InterPro; IPR003137; PA.
InterPro; IPR000209; Peptidase_S8.
Pfam; PF00746; Gram_pos_anchor; 1.
Pfam; PF02225; PA; 1.
Pfam; PF00082; Peptidase_S8; 1.
PRINTS; PR00723; SUBTILISIN.
TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
PROSITE; PS00136; SUBTILASE_ASP; 1.
PROSITE; PS00137; SUBTILASE_HIS; 1.
PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
Signal; Plasmid.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1870
FT PROPEP 1871 1902 PIII-TYPE PROTEINASE.
FT ACT_SITE 217 217 REMOVED BY SORTASE (POTENTIAL).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 1867 1871 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1870 1870 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;

Query Match 6.4%; Score 257.5; DB 1; Length 1902;

```

```

Best Local Similarity   21.0%; Prec. No. 2.3e-09;
Matches 162; Conservative 124; Mismatches 287; Indels 197; Gaps 38;

QY      17 SSSSEILQKQTYYIQLHPNSSETAKTFASKFDFHLSFLQEAVLGVVEEBEESPSSLRLSYG 76
       115 SSTAEIIQQ-----ETNKVIAAQ-----ASVKAAVEQVTQQTAGE---SYG 151
QY      77 SAIEGFAAQITSEAEAILRYSPEVVAVRPDHVLQVQTTYSYKYFLGLDGFNGS-----GVW 131
       152 YVNVGFSTKKRVVDIPKLK-- --QIAGVK-----TWTAKVYYPTDAKANSMANVAQVW 201
QY      132 SKSRP-CQGTHIGVLDGVWPESPSF---DDTGM---PSIPRKWKGTICQEGSESFSSSCN 184
       202 SNYKYKEGTVSVVIDSGIDPTHKDMRLSDDKOVKLTKSDYEKFTDTVHGGRYENSK--- 255
QY      185 RKLIIGASFFIRGHRVANSPBESPNMREYTSARDSTGHGHTATSTVGCGSSVSVMANVLGN 244
       259 ---VPYCF-----NYADNNDTIIDDKVDEQ---HGMHVAGIIAAN-----GTG 295
QY      245 AGVAR---GMAPGAHIAIVKYCWFNG-----CYSSDILAAIDVAIQKDVLDSLGLGF 295
       296 DDPAKSVVGVAPEAQAQLAMKV-- -FNSDTSAKTGSAATVVSIAEDSAKIGADVLLNMSLGN 353
QY      296 P--IPLYDDTIAICTPRAMERGISVICCAANGNPRIESVANAPWV-----STIGAGTL 347
       354 SGNQTLDEPDLA-AVQNANESGTAAVISAGNSGTSGSATGVKNKYDYGLQDNEMVGSFGT 412
QY      348 DRFPFVAVRLANGKLL-YGSLXPKGKIKNAGREV- -----LYVTGGDKGS 393
       413 SRGATTVASAENTDVITQAVTTDGTQLGPETIQJSSHDFQSFQPKKFYIVVDASGN 472
QY      394 EFCIRGLSP--RBIEIRKMVICDRGVNVRSEKGEAVKEAGVAMILANTEINOFEDSIDV 451
       473 --LSKGALADYTADAKGKIALVKEGESFDCKQKYAOAGAAGLIIVNT-----DG 521
QY      452 HLPATLIGVT-----BSULLKAVN-ATVKPKARIIFGGTVIGRSRAP-----E 495
       522 TATPMTSIALTTTPPTFGLSVTCQKLVDMVWTAHPDDSL-- --GVKITAMLPLNQKYTEDK 578
QY      496 VAQESARGPSLANPSILKPDMIAPGVNI IAAVQNLGPTGLPYDSRRVNWTVSGTSMSC 555
       579 MSDFTSYGP-VSNLS-FKPDIAPGGNIWSTQNNN-----GYTNMSGTSMAS 623
QY      556 PHYVSGITAIRSAYPNWSPAARKSALMTTADLYDRQ--GKAIKDGNKPAQGVFAIGAGHYN 613
       624 PFTAGSQALLKQALNNXK-----NPFYAYYKQLKGHTALDFTLKTVE----- 664
QY      614 PQKAINPGLVYNTQPVDYITYTLCTLGFTFRSDILAITHKNVSCNGILRNKPFSL---NYP 670
       665 -----MNTACPIINDINNVIIVSPRQAGLVDRVKAIDA-LEKNFSTVVAENGYP 714
QY      671 SIAY-----IFK-----ROKTMETTRRVNVGPSNISYIVSNVKAPEGI 709
       715 AVSLKQFTSDTKYFKLTFTNRTHLIFYQDMSNTDNTNAVYTSADPNMSGV 764

```

RESULT 9

SUBT_BACSU	STANDARD;	PRT;	361 AA.
ID	SUBT_BACSU		
AC	P04189; C07613; P70989;		
DT	20-MAR-1987 (Rel. 04, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Subtilisin E precursor (EC 3.4.21.62).		
GN	APRE OR APRA OR SPRE.		
OS	Bacillus subtilis.		
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
OX	NCBI_TaxID=1423;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP			
RC	STRAIN=168;		
RC	MEDLINE=84212198; PubMed=6427178;		
RA	Stahl M.L., Ferrari E.;		

RT "Replacement of the *Bacillus subtilis* subtilisin structural gene with
RT an in vitro-derived deletion mutation.";
RL J. Bacteriol. 158:411-418(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98240224; PubMed=9579061;
RA Noback M.A., Holsappel S., Kiewiet R., Terpstra P., Wambutt R.,
RA Wedler H., Venema G., Bron S.;
RT "The 172 kb *prfA*-*addA* region from 83 degrees to 97 degrees of the
RT *Bacillus subtilis* chromosome contains several dysfunctional genes,
RT the *glyB* marker, many genes encoding transporter proteins, and the
RT ubiquitous *hit* gene.";
RL Microbiology 144:859-875(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9394377;
RA Avesto F., Ogawara N., Moszer I., Albertini A.M., Alloni G.,
RA Kuneš V., Bertero M.G., Bessières P., Bolotin A., Borchert S.,
RA Borriès R., Boursier L., Brans A., Braun M., Brighelli S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Daniot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert B., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Jorba B., Karamat D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinio S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Fohl T.M., Portetelle D., Rapoport S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rappollik S., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler H., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumslein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256(1997).
RN [4]
RP SEQUENCE OF 1-156 FROM N.A.
RC STRAIN=168 / PY79;
RX MEDLINE=8414862; PubMed=6322190;
RA Wong S.L., Price C.W., Goldfarb D.S., Doi R.H.;
RT "The subtilisin E gene of *Bacillus subtilis* is transcribed from a
RT sigma 37 promoter in vivo.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:1184-1188(1984).
RN [5]
RP SEQUENCE OF 1-156 FROM N.A.
RC STRAIN=168 / PY79;
RX MEDLINE=87222417; PubMed=3108260;
RA Ikemura H., Takagi H., Inouye M.;
RT "Requirement of pro-sequence for the production of active subtilisin E
RT in *Escherichia coli*.";
RL J. Biol. Chem. 262:7859-7864(1987).
RN [6]
RP SEQUENCE OF 1-13 FROM N.A.
RC STRAIN=168;
RX MEDLINE=89213955; PubMed=2436113;
RA Park S.S., Wong S.L., Wang L.F., Doi R.H.;
RT "Bacillus subtilis subtilisin gene (*aprE*) is expressed from a sigma A
RT (sigma 43) promoter in vitro and in vivo.";
RL J. Bacteriol. 171:2657-2665(1989).
RN [7]

RN SEQUENCE OF 1-8 FROM N.A.
 RP MEDLINE=89868685; PubMed=2447063;
 RX Hemmer D.J.; Ferrarri E.; Perego M.; Hoch J.A.;
 RA "location of the targets of the hpr-97, sacU32(Hy), and sacQ36(Hy)
 RT mutations in upstream regions of the subtilisin promoter.";
 RL J. Bacteriol. 170:296-300(1988).
 [8]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RC STRAIN=168;
 RX MEDLINE=99030466; PubMed=9811547;
 RA Jain S.C.; Shinde U.; Li Y.; Inouye M.; Berman H.M.;
 RT "The crystal structure of an autoprocessed Ser221Cys-subtilisin
 RT E-propeptide complex at 2.0-A resolution.";
 RL J. Mol. Biol. 284:137-144(1998).
 CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE.
 CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 CC for peptide bonds, and a preference for a large uncharged residue
 CC in fl. Hydrolyzes peptide amides.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
 CC SPORELLATION, AND MANY MUTATIONS WHICH BLOCK SPORELLATION AT EARLY
 CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
 CC IS NOT NECESSARY FOR NORMAL SPORELLATION.
 CC -!- SIMILARITY: Belongs to peptidase family S8.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; K01988; AAA22742.1; -;
 DR EMBL; Y14083; CAA74536.1; -;
 DR EMBL; Z99109; CAB12870.1; -;
 DR EMBL; K01443; AAA22814.1; -;
 DR EMBL; M16639; AAA22744.1; -;
 DR EMBL; M31060; AAA22246.1; -;
 DR EMBL; M19125; AAA22245.1; -;
 DR PIR; A00972; SUBS1.
 DR PDB; 1SCU; 13-JAN-99.
 DR MEROPS; S08.036; -;
 DR Subtilist; BG10190; aprE.
 DR InterPro; IPR000209; Peptidase S8.
 DR Pfam; PF00082; Peptidase S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 KW Hydrolase; Sporulation; Serine protease; Zymogen; signal;
 KW 3D-structure; Complete proteome.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 106 POTENTIAL.
 FT CHAIN 107 381 SUBTILISIN E.
 FT ACT_SITE 138 138 CHARGE RELAY SYSTEM.
 FT ACT_SITE 170 170 CHARGE RELAY SYSTEM.
 FT ACT_SITE 327 327 CHARGE RELAY SYSTEM.
 FT CONFLICT 27 27 V -> A (IN REF. 1, 4 AND 5).
 FT STRAND 37 43
 FT HELIX 52 60
 FT TURN 61 63
 FT STRAND 65 69
 FT STRAND 75 80
 FT HELIX 82 89
 FT TURN 90 90
 FT TURN 92 93
 FT STRAND 94 99
 FT STRAND 102 105
 FT HELIX 112 116
 FT TURN 117 118

FT HELIX 119 125
 FT TURN 126 131
 FT TURN 133 138
 FT STRAND 143 144
 FT TURN 146 147
 FT STRAND 150 155
 FT TURN 158 159
 FT TURN 163 164
 FT TURN 170 179
 FT TURN 192 193
 FT STRAND 195 200
 FT TURN 204 205
 FT STRAND 207 209
 FT STRAND 210 222
 FT TURN 223 224
 FT STRAND 227 230
 FT STRAND 232 234
 FT HELIX 239 250
 FT TURN 251 252
 FT STRAND 254 258
 FT STRAND 265 265
 FT TURN 266 267
 FT STRAND 268 268
 FT STRAND 273 273
 FT TURN 274 276
 FT TURN 278 279
 FT STRAND 281 286
 FT TURN 288 289
 FT STRAND 292 292
 FT TURN 294 295
 FT TURN 300 301
 FT STRAND 304 307
 FT STRAND 311 315
 FT TURN 316 318
 FT STRAND 319 323
 FT HELIX 326 343
 FT TURN 345 346
 FT HELIX 349 358

 QY 6 PFLCIIFLLFCSSSSSEILQQTIVVQLHPNSETAKTFASKFDWHLSPLOEAVLGVSEEE 65
 DB 12 FALTLIIFTMAFSNMS-----VQAAGKSTSEKKYIVGFKQTMSSAMSAKKKDVISEK 62
 QY 66 EPSRRLLYSYGSAIEGFAAOLTESEAILRYSPEWAVRDPDVLQVQTTYSYKFLGIDGF 125
 DB 63 GGRVKQKQFKY---VNAAAATLDEKAVKELKDFSAIVEEDHIAH-EYAGSVFY-GISQI 117
 QY 126 GNSGVMSKSRPQGQTTIIGVLDTGWVPSPSPFDDTGMPSIPRKWKGICQEGESFSSSCNR 185
 DB 118 KAPALHSQGTGSNVKVAVIDSGIDSSHPLNVRG----- 152
 QY 186 KLICARFFIRGHVRVANSPESSPNMPREYISARSTGHGTHTASTVGSSSVSMANVLNGA 245
 DB 153 ---GASF-----VPSETNPYQGSSSHGTHVAGTIA-----ALNNSI 185
 QY 246 GVAEGMAPGAHIAVYKVCWFNGC-YSSDILLAAIDVAIQDKVDVLSLGLGFPPIPLYDDTI 304
 DB 186 GVL-GVSPFASLYAVKVLDDSTGSGQYSWIINGIEWAISSNNDVINMSLGG---PTGSTAL 241
 QY 305 AIGTFRAMEGISVICAAAGNNGPIESSVANTAPWVSTIGAGTILDRRFPVAVRLANGKLLY 364
 DB 242 KTVVDKAVSSGIVVAAAAGNEGSSGST-----STVG----- 272
 QY 365 GESLYPGKGIKNAGREVEVIYVTGDKGSEFCLRGSLPREIRGNVICDRGVNGRSEKG 424
 DB 273 ---YPAK----- 276
 QY 425 EAVKEAGGVAMILANTEINQEEDSIDVHLLPATL-IGYTESVLLKAYVNAVTKPKARIIF 483

Query Match 6.4%; Score 257; DB 1; Length 381;
 Best Local Similarity 21.0%; Pred. No. 2.9e-10;
 Matches 129; Conservative 62; Mismatches 175; Indels 248; Gaps 20;

```

Db 277 -----YPSIIAGVANS----- 288
Qy 484 GGTVIGRSRAPEVAQFSAQPSLANPSILKPDMPAIPCVNIIAAWPNQGLPTGLPYDSRRV 543
Db 289 -----SNQRASTFSSAGSEL-----DWAAGVSI-----QSTLPGG----- 318
Qy 544 NFTVMSGTSMSCHVSGITILIRSVNPNWSPAIAKALMTADLYDRQKAIKDGKNKAG 603
Db 319 TYGAYNGTSMATPHVAGAAAILSKIPTWNAQVRDLSESTATYL-----GNS----- 366
Qy 604 VFAIGAGHVNPOKA 617
Db 367 -FYGKGLINVOAA 379

RESULT 10
P1P_LACLC
ID -P1P_LACLC STANDARD; PRT; 1902 AA.
AC P16271;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE P1-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
DE proteinase).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pW05.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN SEQUENCE FROM N.A.
RC STRAIN=WG2;
RX MEDLINE=98149035; PubMed=3278687;
RA Kok J., Leenhouks K.J., Haandrikman A.J., Ledebor A.M., Venema G.;
RT "Nucleotide sequence of the cell wall proteinase gene of
RT Streptococcus cremoris Wg2.";
RL Appl. Environ. Microbiol. 54:231-238(1988).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC -!- GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC Pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyze hemoglobin and oxidized
CC insulin B-chain.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24767; AAA17677.1; -
CC DR HSSP; P00782; 1501.
CC DR MEROPS; S08.019; -.
CC DR InterPro; IPR001899; Gram_pos_anchor.
CC DR InterPro; IPR006192; LPXTG.
CC DR InterPro; IPR003137; PA.
CC DR InterPro; IPR000209; Peptidase_S8.
CC DR Pfam; PF00746; Gram_pos_anchor_1.
CC DR Pfam; PF02225; PA; 1.
CC DR Pfam; PF00082; Peptidase_S8; 1.
CC DR PRINTS; PR00723; SUBTILISIN.
CC DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
CC DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
CC DR PROSITE; PS00136; SUBTILASE_ASP; 1.
CC DR PROSITE; PS00137; SUBTILASE_HIS; 1.

```

```

DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
KW SIGNAL; Plasmid.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1870
FT PROPEP 1871 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT SITE 1867 1871
FT MOD_RES 1870 1870
SQ SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5D0B CRC64;

Query Match
Best Local Similarity 21.9%; Pred. No. 2.5e-09;
Matches 173; Conservative 115; Mismatches 285; Indels 218; Gaps 40;

QY 17 SSSSEILQKQYIVQLHPNSETAKTFASKFDWHLFSLQEAFLGVEEBEESPSSLLYSYG 76
Db 115 SSTAELIQ-----ETNKVIAAQ-----ASVKAAVEQVTQQTAGE---SYG 151
QY 77 SAIGFPAQLTESAEAILRYSPEVAVRPHVLQVQTYSYKFLGLDGFNS-----GVW 131
Db 152 YVANGFSTKRVVDIPKLK---QIAGVK-----TVTLAKVYPTDAKANSMANVQAV 201
QY 132 SKSRF-GQGTIIIGVLDTGWMPESPSP---DTGM---PSIPRWKKGICQGESFSSSCN 184
Db 202 SNYKKEGTIVSVSDISGIDTHDKMLSDDKVKLTKSDVERKFTDTAKHGRYNSK--- 258
QY 185 RKLIGARFFIRGHRVANSPEESPMPREYISARDST---GHGHTASTVGSGSSVSMANVL 241
Db 259 -----VPYGFNVADND-----TITDDTVDEQGHMVAGIIGAN----- 292
QY 242 GNGAGVAR---GMAPGAHIAVYKWFNG-----CYSSDILAAIDVAIQKVDVLSLSLG 293
Db 293 GTGDDPAKSVVGVAPAEQLAMKV-FTNSDTSATGSSTLVSALEDSAKIGADVLNMSLG 351
QY 294 GFP--IPLYDDTIAIGTFRAMEGISVICAGNNGPIESS-----VA 333
Db 352 SDSGNQTLDEPDLA-AVQNANESGTAAVISAGNSGTSATGEGVKNQDYGLQDNEMVGT 410
QY 334 NTAPWVSTIGAGTLDRAFPAPVRLANGKLLYGEISLYPKGKIKNAGREV-----EVIY 387
Db 411 GTSRGATTASAENTDVITQAVTIDGT---GLQLGPGT-IQLSSNDFTSGFDQKKFVV 466
QY 388 GGDGSEFCLRGSLP--REEIRGRV-CDRGVNGRSEKGEAVKBAAGVAMILANTEINQE 445
Db 467 KDAAGN--LSKDALADYTADAKGKIAIVKRGELSFDKQKYAQAAGAGLIIVNN----- 519
QY 446 EDSIDVHLLPATLIGYT-----ESVLLKAYVN-ATVVKPARIIFGGTVIGSRAP-- 494
Db 520 ---DGTATPVTSMALTTTFTFGLSSVTGQKLDVWVTAHPDSDL---GVKIALTLVPNQ 572
QY 495 -----EVAQFSARGPSLANPSILKPDMPAIPCVNIIAAWPNQGLPTGLPYDSRRVNFVMS 549
Db 573 KYTEKMSDFTSYGP-VSNLS-FKPDITAPGNIWSTQNNN-----GYTNMS 617
QY 550 GTSMSCHVSGITILIRSVNPNWSPAIAKALMTADLYDRQKAIKDGKNKAGS--- 589
Db 618 GTSMASPFIAQSQAALLKQALNNKNNPFYAYYKQKLTALTFDLKTVEMTAQPINNDYN 677
QY 590 -----RQKAIKDGKNKAGVFAIGAGHVNPQKAIKALNPGVLYIQPVYDITYLCT--LGF 640
Db 678 NVIVSPRQAGLVDVKA---AIDALEKNPSTVAENGYPVAVELKDFSTDKTFLTF 732
QY 641 TRSDILAITHK---NVSCNGILRKNGPFSLNYSIAVIFKRGKTEMITRVTNNGSPNS 697
Db 733 TNSITHELTQOMDSNTDTNAVYT-----SATDPNSGLYDK-----KIDGAIAKAGS--- 779
QY 698 IYSVNVKAPEG 708
Db 780 ----NITVPAG 786

```

```

RESULT 11
SUBT_BACSA
ID SUBT_BACSA STANDARD; PRT; 381 AA.
DT P00783;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Subtilisin amylosacchariticus precursor (EC 3.4.21.62).
GN APR.
OS Bacillus subtilis var. amylosacchariticus.
OC Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1483;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89008194; PubMed=3139650;
RA Yoshimoto T., Oyama H., Honda T., Tone H., Takeshita T.,
RA Kamiyama T., Tsuru D.;
RT "Cloning and expression of subtilisin amylosacchariticus gene.";
RL J. Biochem. 103:1060-1065(1988).
RN [2]
RN PARTIAL SEQUENCE.
RX MEDLINE=72266687; PubMed=4560201;
RA Markland F.S., Kurihara M., Smith E.L.;
RT "Subtilisin Amylosacchariticus. II. Isolation and sequence of the
RT tryptic and cyanogen bromide peptides.";
RL J. Biol. Chem. 247:5602-5618(1972).
RN [3]
RN SEQUENCE OF 107-381.
RX MEDLINE=72266688; PubMed=5055784;
RA Kurihara M., Markland F.S., Smith E.L.;
RT "Subtilisin Amylosacchariticus. 3. Isolation and sequence of the
RT chymotryptic peptides and the complete amino acid sequence.";
RL J. Biol. Chem. 247:5619-5631(1972).
CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
CC for peptide bonds, and a preference for a large uncharged residue
CC in P1. Hydrolyzes peptide amides.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
CC SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
CC IS NOT NECESSARY FOR NORMAL SPOULATION.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00264; BAA00186.1; -.
DR PIR; A41448; SUBSS.
DR HSP; P04189; 1SCJ.
DR MEROPS; S08.042; -.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Hydrolase; Sporeulysin; Serine protease; Zymogen; Signal.
DR SIGNAL
DR PROPEP 1 30
DR POTENTIAL.
DR PROPEP 31 106
DR FT CHAIN 107 381 SUBTILISIN AMYLOSACCHARITICUS.
DR FT ACT_SITE 138 138 CHARGE RELAY SYSTEM (BY SIMILARITY).
DR FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (BY SIMILARITY).
DR FT ACT_SITE 327 327 CHARGE RELAY SYSTEM (BY SIMILARITY).
DR FT CONFLICT 191 191 S -> A (IN REF. 2).

```



```
FT HELIX 117 120
FT TURN 121 123
FT HELIX 125 130
FT TURN 131 131
FT TURN 135 136
FT TURN 138 142
FT STRAND 150 151
FT TURN 156 157
FT STRAND 173 182
FT STRAND 198 201
FT TURN 207 208
FT HELIX 213 226
FT STRAND 230 233
FT HELIX 242 252
FT TURN 253 255
FT STRAND 257 258
FT TURN 267 268
FT TURN 272 275
FT STRAND 279 279
FT STRAND 282 285
FT STRAND 291 291
FT TURN 299 300
FT STRAND 303 306
FT STRAND 310 314
FT TURN 315 317
FT STRAND 318 322
FT HELIX 325 342
FT TURN 344 345
FT HELIX 348 358
FT HELIX 375 380
SQ SEQUENCE 380 AA; 539EA7271B6682C CRC64;

Query Match 6.2%; Score 248; DB 1; Length 380;
Best Local Similarity 19.9%; Pred. No. 1.2e-09;
Matches 123; Conservative 68; Mismatches 164; Indels 262; Gaps 22;

QY 10 IIFLLCSSSEILQ--KQYIVQLHPNSETAKTFASKDFDHLSPFQEAFLGVVEEEEP 67
DB 15 LISVASSSSASAEAEKEXLIGFN-EQAVSEFVEQE-----ANDEVAILSEEEV 67

QY 68 SSRLLSYGSAIEGFAAQLTESAEILRYSPVEYAVRPHVLQVQTYGYKFLGDGFN 127
DB 68 EIELLHEF-ETIPVLSVELSPEDVDALDPAISYIEED--AEVTTMAQSVPMGSRVQA 124

QY 128 SGWWSKRFQGGIIGVLTGWBPSPSPDDTGMPSIPRWKIGICOGESFSSSSCNRL 187
DB 125 PAAHNRGLTSGVKVAVLDTGI-----STHPDLNI 154

QY 188 IGARFFTRGHRVANSPEESPMPREVISARDSTGHGTHTASTVGGSSVSMANVLNGAGV 247
DB 155 RGGASFV-----PGEP-----STQDGNHGHGVAGTIA-----ALANNSIGV 190

QY 248 ARGMAPGAHIAVYKVCWFNGCYS-SDILAAIDVAIQDKVDVLSLSLGGFPPIPLYDDTTAI 306
DB 191 L-GVAPNAELIYAVKVLGASGSGSVSIAQGLEWAGNNGMHVANSLSGSPSPATLEQ 246

QY 307 GTFRAMEGTSVTCAGNNGPIESSVANTAPWSTTGAGTLDRFPFVAVLANGLLYGE 366
DB 247 AVNSATSRGVLVVAASGNS-----GAGSIS--YPA--RYAN----- 278

QY 367 SLIPGKIKNAGEVEVIYTGDKGSEPCLRGSLPREIRGNWVICDRGVNRSKGEA 426
DB 279 ----- 278

QY 427 VKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYVNATVKPKARIIFGCT 486
DB 279 -----AMAVGATDQNNR----- 291

QY 487 VIGSRAPVQAQPSARGPSLANPSILKPDWIAFGVNNIIAWPONLGTGLPYDSRRVNFT 546
DB 292 -----ASFQYAGL-----DIVAEGVNVQSTYP-----GSTVAS----- 321

QY 547 VMSGTNSCPHVSIGITALIRSAYPNWSPAIKSALMTTA-----DLYDRQKAIKDGKNK 600
```

```
DB 322 -LNGTSMATPHVAGAAALVQKPNPSNSVQIRNHLKNTATSLGSTNLY----- 368
QY 601 PAGVFAJGAGHVNPOKA 617
DB 369 -----GSLVNAEAA 378

RESULT 15
ELVA_BACCS STANDARD; PRT; 380 AA.
AC P41362;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alkaline protease precursor (EC 3.4.21.-).
OS Bacillus clausii.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=79880;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
RX MEDLINE=93043753; PubMed=1368952;
RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakamura S.,
RA Aono R., Horikoshi K.;
RT "Molecular cloning, nucleotide sequence, and expression of the
RT structural gene for alkaline serine protease from alkaliphilic
RT Bacillus sp. 221.";
RL Biosci. Biotechnol. Biochem. 56:1455-1460(1992).
RN [2]
RP SEQUENCE OF 112-129.
RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
RA Horikoshi K.;
RL (In) Horikoshi K. (eds.);
RL Microorganisms in alkaline environments, pp.187-194, VCH,
RL Weinheim (1991).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; S48754; AAC60420.1; -
CC EMBL; D13157; BAA02442.1; -
CC EMBL; A26817; CAA01836.1; -
CC EMBL; A22550; CAA01611.1; -
CC HSSP; P29600; 1GCI.
CC MEROPS; S08.103; -
CC InterPro; IPR000209; Peptidase S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE ASP; 1.
CC PROSITE; PS00137; SUBTILASE HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 111
FT CHAIN 112 380
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 326 326 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 380 AA; 5F73ABC6D5B6831 CRC64;

Query Match 6.2%; Score 248; DB 1; Length 380;
Best Local Similarity 19.9%; Pred. No. 1.2e-09;
Matches 123; Conservative 68; Mismatches 164; Indels 262; Gaps 22;

QY 10 IIFLLCSSSEILQ--KQYIVQLHPNSETAKTFASKDFDHLSPFQEAFLGVVEEEEP 67
```

Db 15 LISVAFSSSIAGAAEKEKYLIGN-EQEAUSEFEQVE-----ANDEVAILSEEEV 67
QY 68 SSRLLYSYGSATIEGFAAQTESEAILRYSPEVAVRDPHVLQVOTTYSYKFLGLDGFN 127
Db 68 ETELLHEF-ETIPVLSVELSPEDVDALEIPAIYIEED--AEVTTMAQSPVPGISRYQA 124
QY 128 SGVWSKSRFGQGTIGVLDTGVPWSPSPDDTGMPISPRKWKIGICQEGESPSSSCNPKL 187
Db 125 PAHNRGLTSGVKVAVLDTGI-----STHPDLNI 154
QY 188 IGARFFIRGHRVANGSPPEPNPREYISARDSTGHGTHTASTVGGSSVSMANVLGNAGV 247
Db 155 RRGSEFV-----PGEP-----STQDNGHGHVAGTIA-----ALNNSIGV 190
QY 248 ARGMAPGAHIYKVCWFNGCYS-SDILAAIDVAIQDKVDLSLSLGFPPIPLYDDTIAI 306
Db 191 L-GVAPSAELYAVKVLGAGSGSVSSIAQGLEWAGNMGHMANLSLGS---PSPSATLEQ 246
QY 307 GTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGAGTIDRRPFAVRLANGKLLYGE 366
Db 247 AVNSATSRGLVVAASGNS-----GAGSIS--YEA--RYAN----- 278
QY 367 SLYPGKGIKNAGREVEVIVTGGDKGSEFCLRGLPREIRGKMWICDRGVNGRSEKGEA 426
Db 279 ----- 278
QY 427 VKEAGGVAMILANTEINQEEEDSIDVHLLPATLIGYTESVLLKAYVNAIVKPKARIIFGOT 486
Db 279 -----AMAVGATDQNNR----- 291
QY 487 VIGRSRAPEVAQPSARGPSLIAPSIKPDMIAPGUNIIAAMPONLPGTCLPYDSRRVNET 546
Db 292 -----ASFSQYAGL-----DIVAPGVNVQSTYP-----GSTYAS----- 321
QY 547 VMSGTWSGPHVSGITALIRSAVPNWSPAIKSALMTTA-----DLYDROGKAIDGNK 600
Db 322 -LNGTSMATPHVAGAAALVKQKNPSWSNVQIRNHLKNTATSLGSTNLY----- 368
QY 601 PAGVFAIGAGHVNPOKA 617
Db 369 -----GSGLVNAEAA 378

Search completed: February 2, 2004, 09:10:32
Job time : 21 secs

SQ SEQUENCE 775 AA; 83775 MW; 7740B20397C7C211 CRC64;

Query Match 100.0%; Score 4018; DB 10; Length 775;
 Best Local Similarity 100.0%; Pred. No. 2.5e-265;
 Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKPPFFLCIIIFLLFCSSSEILQKQYIVOLHNSTAKTFASKDWHLSFLQEAFLGV 60
 DB 1 MEKPPFFLCIIIFLLFCSSSEILQKQYIVOLHNSTAKTFASKDWHLSFLQEAFLGV 60

QY 61 EEEBEPSSRLLYSGSAIEGFAAQLTESEAEILRYSPVAVRPHVLOVQTTYSKPL 120
 DB 61 EEEBEPSSRLLYSGSAIEGFAAQLTESEAEILRYSPVAVRPHVLOVQTTYSKPL 120

QY 121 GLDGFNSGVWKSFGGTTIGVLDGVWPESSEDDTCMPISIPRKWKGIQEGESFSS 180
 DB 121 GLDGFNSGVWKSFGGTTIGVLDGVWPESSEDDTCMPISIPRKWKGIQEGESFSS 180

QY 181 SSCNRKLGARFFIRHGVANSPESPNMPREYISARDSTGCHTASTVGGSSVMANV 240
 DB 181 SSCNRKLGARFFIRHGVANSPESPNMPREYISARDSTGCHTASTVGGSSVMANV 240

QY 241 LKNGAGVARGMAPAGHIAVYKVFNGCYSDILAAIDVAIQKVDVLSLGGPPIPL 300
 DB 241 LKNGAGVARGMAPAGHIAVYKVFNGCYSDILAAIDVAIQKVDVLSLGGPPIPL 300

QY 301 DDTIAGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTLDRRFAVVRLANG 360
 DB 301 DDTIAGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTLDRRFAVVRLANG 360

QY 361 KLYGESLYPGKIKNAGREVEVYVYVGGKSGFECLRGSLPREETRGKMWICDRGVNR 420
 DB 361 KLYGESLYPGKIKNAGREVEVYVYVGGKSGFECLRGSLPREETRGKMWICDRGVNR 420

QY 421 SEKGAIVKAGVAMILANTINOEDSIDVHLLPATLIGYTESVLLKAYVNAVTKPKAR 480
 DB 421 SEKGAIVKAGVAMILANTINOEDSIDVHLLPATLIGYTESVLLKAYVNAVTKPKAR 480

QY 481 IIFGTVIGRSAPVAOFSARGPSLANPSILKPDMIAPGVNTIAWPNQGLPYDS 540
 DB 481 IIFGTVIGRSAPVAOFSARGPSLANPSILKPDMIAPGVNTIAWPNQGLPYDS 540

QY 541 RVNFTVMSGTSMSCPHVSIGITALIRSAYPNWSAAIKSALMTTADLYDROGKAKDGNK 600
 DB 541 RVNFTVMSGTSMSCPHVSIGITALIRSAYPNWSAAIKSALMTTADLYDROGKAKDGNK 600

QY 601 PAGVFAIGAGHVPKAINPGLVNIOPVDYITVCTLGTFRSDILAITHKNVSCNGILR 660
 DB 601 PAGVFAIGAGHVPKAINPGLVNIOPVDYITVCTLGTFRSDILAITHKNVSCNGILR 660

QY 661 KNPGLSLNYPSTAVIFKKGKTEMITRVTNNGSPNSIYVNVKAPGKIVNPKRLVF 720
 DB 661 KNPGLSLNYPSTAVIFKKGKTEMITRVTNNGSPNSIYVNVKAPGKIVNPKRLVF 720

QY 721 KHVDTLSYRVNVLKKNRGKVASPAQGLTWNHNLQVRSPISVTLKTN 775
 DB 721 KHVDTLSYRVNVLKKNRGKVASPAQGLTWNHNLQVRSPISVTLKTN 775

RESULT 2
 Q9ZUF6 PRELIMINARY; PRT; 754 AA.
 AC Q9ZUF6;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative serine protease (Putative subtilisin serine protease).
 GN AT2G05920.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 CX NCBI_TaxID=3702;

[1] SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White C., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:761-768(1999).
 [2] SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RA Lin X.;
 [3] SEQUENCE FROM N.A.
 RP Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
 RA Bowers L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
 RA Jones T., Kamiya A., Karir-Neumann G., Kawai J., Kim C., Koesema E.,
 RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
 RA Palm C.J., Sakurai T., Satou M., Seki M., Shim P., Southwick A.,
 RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene T6P5.12/At2G05920 (GI:4006827).";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 [4] SEQUENCE FROM N.A.
 RP Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,
 RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C.,
 RA Wu H.C., Yu G., Yuan S., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Tripp M.G., Wu T., Shinzaki K., Davis R.W.,
 RA Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; ACC05970; AAC95169.1; -
 DR EMBL; AY035090; AAK9595.1; -
 DR EMBL; AY142613; AANJ3182.1; -
 DR HSP; Q99405; IMPT; -
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 2.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00840; PA; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Protease.
 SQ SEQUENCE 754 AA; 80015 MW; P8CCG3F1L578CEP7 CRC64;

Query Match 43.4%; Score 1742; DB 10; Length 754;
 Best Local Similarity 48.4%; Pred. No. 3.9e-110;
 Matches 370; Conservative 118; Mismatches 237; Indels 40; Gaps 16;

QY 11 IFLLFCSSSEILQKQYIVOLHNSTAKTFASKDWHLSFLQEAFLGVVEEPESSR 70
 DB 18 LFLLETTA-----KKTIVIRVN-HSDKPSFULTHHDWYTSQLSN-----ESS 59

QY 71 LLYVSGSAIEGFAAQLTESEAE-ILRYSPVAVRPHVLOVQTTYSKFLGLDG-FGNS 128
 DB 60 LLYVTTTFHGFSAVLDSTEDSLSSNSILDFEDPLTYLHTTTPPEPLGNSFGVH 119

QY 129 GWMSKSRFGQTTIGVLDGTGVWPSPFDGTGMPSPRKGKICQEGESFSSSCNRKLI 188
 DB 120 DLGSSS---NGVIIGVLDGTGVWPSPRFDGTDMPSPRKGKICQEGESFSSSCNRKLI 176

QY 189 GARFTIRGHRVANSPEESPNNPREYISARDSTGCHTASTVGGSSVMANVLCNGAGVA 248
 DB 189 GARFTIRGHRVANSPEESPNNPREYISARDSTGCHTASTVGGSSVMANVLCNGAGVA 248

Db 177 GARSFSGFQMSAGCGFSK--RESVPRDVGHGHTTTTAAGSAVRNASFLGVAAGTA 234
QY 249 RGMAGAHIAVXKWFNGCYSSDILAAIDVAIQKVDVLSLGGFPIPLDYDDTIAICT 308
Db 235 RGMATRATVATYKVCSTGFCGSDILAAANDRAILDGVVLSLGGGSAFYRDIIAGA 294
QY 309 FRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTLDLDRFPVAVRLANGKLLYGSSL 368
Db 295 FSAMERGVSFVSCAGNSGPTASVANVAPWMTVGAGTLDLDRFPAPANLGNKRLTGVS 354
QY 369 YPGKGIKNAGREVEVLYVGGKSEFCLESGIPREIRGKMWICDRGVNGSEGEAVK 428
Db 355 YSGVGM--GTXPLELYNKGSSSNLCUPGSLDSSIVRGKIWCDRGVNAKVEGAVR 412
QY 429 EAGGVAMILANTEINOEDSIDVHLLPATLIGVTESVLKAYNATVVKPARIIEGGTVI 488
Db 413 DAGGLWIANWTAASGEELVANSLLPATAVGKKTGDLLEVVKSDSFTALLVFKGTVL 472
QY 489 GRSRAPEVAQFARGPSLANPSILKPDMIAPGNIIAAMPONLGTGLPYDSRRYNTVM 548
Db 473 DVKPSVWAFAFSRGNTVTEILKPDVIGPGVNLILAGWSDAIGPTGLDKDSRRTOFNIM 532
QY 549 SGTSMSCPHVSGITAIRISAYPNWSPAAIKSALMTADLYDRQKAIKDG--NKPAQVFA 606
Db 533 SGTSMSCPHISGLAGLLKAAHPWSPSAIKSALMTAYVLDNTNAPLHDAADNSLSNFA 592
QY 607 IGAGHYNPOKAINPGLVNIQPDVITYLTCTLGTFRSDILAITHK--NVSCNGLIRKNPGF 665
Db 593 HSGGHVDPQKALSPGLVYDITSTEEYIRFLCSDIDYVDHIVALVKPSVNCCKF--SDFG- 650
QY 666 SLNYPISVIFKRGKTEMITRRVTNVGSPNSIYVNVKAPGKIVIVNPKRLVFKHDQ 725
Db 651 QLNYPFSVLF--GGKRVVRYTREVNNGAASSVYKVTYNGAFSGISVKPSKLSFKSGE 709
QY 726 TUSYRVFVFLKKNRGKGVASFAQ--GQLTVNWSHNLMOVRSPIS 769
Db 710 KRYTTFVSKK---GVSMNKAEFGSITWSNPOH---EVRSPVA 748

RESULT 3

Q9LUM3 ID Q9LUM3 PRELIMINARY; PRT; 775 AA.
AC Q9LUM3
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Subtilisin proteinase-like protein.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Nakamura Y.;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones";
RL DNA Res. 7:131-135(2000).
DR EMBL; AB022220; BAB01030.1; --
DR HSP; Q99405; IMPT.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.

RESULT 4

Q9CSN5 ID Q9CSN5 PRELIMINARY; PRT; 775 AA.
AC Q9CSN5
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 82.5 kDa protein.

DR PROSITE; PS50840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 775 AA; 82584 MW; CC909DB4ADABDE08 CRC64;
Query Match 42.3%; Score 1700; DB 10; Length 775;
Best Local Similarity 45.0%; Pred. No. 3e-107;
Matches 356; Conservative 134; Mismatches 257; Indels 44; Gaps 15;
QY 6 FFLCIIFLLSCSSSETLQK---TVIQLHPNSETAKTFASKFDWHLSPQLQEAVLGVEE 62
Db 3 FFYFFFLLLTSFSSASSSSNSLTIVVHDHAKPS-IPPTHFWYTSLSLIT----- 56
QY 63 EEEPPSRLLYSXGAIEGFAAQLTSEAEILRYSPEWAVRPHVLQVQTTYSYKFLGL 122
Db 57 --SSPPS-IITHYDVTDFHGFSAELTSCDASQLLDHPHVISIPEQVRHLHTTSPFLGL 113
QY 123 DFGNSGVWSKSRFGQGTIIIGVLDTGWPSPSFDGTGMPISIRKWKGIQCEGESSSSS 182
Db 114 RSTDKAGLLSEDFGSLVIGDVTGWVPERPFDRLGPGVPIKWKGCQIASQDFESA 173
QY 183 CNRKLIGARFFIRGHRVANSPEESPNMPREYIARDSTGHGHTASTVGGSSYSMAVNLG 242
Db 174 CNRKLVGARFFCGGYEATNG---XNNETTEFRSDSDGHGHTASISAGRYVFPASTLG 230
QY 243 NGAGVARGMAPAGAHIAVYKVCWFNGCYSSDILAAIDVAIQKVDVLSLGGFPIPLYDD 302
Db 231 YAHGVAAGMAPAKARLAAYKVCWNSGCYDSIDILAAFDATAVDGVDVLSLGGVVPYLD 290
QY 303 TIAIGTFRAMEGISVICAAGNNGPIESSVANTAPWSTIGAGTLDLDRFPVAVRLANGKL 362
Db 291 AIAIGAFGAIDRGIFFVSASAGNGGPGALTIVTNVAPWMTTVGAGTIDRDPANVKLGNGKM 350
QY 363 LYGESLYPGKGIKNAGREVEVIY---VTGDD-KGSFCLRGSLPREIRGKMWICDRGVN 418
Db 351 ISGVSVYGGPGL--DPCRMVPLVYGGSLGGDGYSSSLCLEGSLDPLNPKVILCDRGIN 409
QY 419 GRSEKGEAVKEAGGVAMILANTEINOEDSIDVHLLPATLIGVTESVLLKAYVNAVTK-- 476
Db 410 SRATKGEIVKNGGLGMIANGVFDGEGLVADCHVLPATSVGASGGEIRRYISESKSR 469
QY 477 ----PKARIIFGGTVIGRSRAPEVAQFARGPSLANPSILKPDMIAPGNIIAAMPONLG 532
Db 470 SSKEPTATVIFKGTGLGIRPAPVVSFARGPNPETEILKPDVIAPIAGNLIAAWPDRIG 529
QY 533 PTGLPYDSRRYNTVMSTGSMSCPHVSGITAIRISAYPNWSPAAIKSALMTADLYDRQG 592
Db 530 PSGYTSDNRETEFNILSGTSMACPHVSGLAALIKAAHPDMSPAAIRSALITATTYVDNSG 589
QY 593 KAIKD--GNKPAGVPAIGAGHYNPOKAINPGLVNIQPDVITYLTCTLGTFRSDILAIT 649
Db 590 EPMMDSTGN--TSSVMDYGGHGHVHTKAMPDGLVYDITSYDIINFLCNSTYTKNTVIT 648
QY 650 HKVSCNGILRKNPGFSLNYPISAVIFKRGKTEMIT---RRVTNVGSPNSIYSVNVKAP 706
Db 649 RRACDCDAGRAGHVGNLNYPFSVSVFQQYGESKMTHTFRTVTNVGSDSVETIKRPP 708
QY 707 EGKIVINPKRLVFKHDQTLISYRVNMFVLKKNRGKGVASFA----QGQLTVNWSHNLMO 762
Db 709 RGTVTVEPEKLSFRVVGQKLS---FVVRVKTEVKLSFGATNVETGHIWSDG---KR 761
QY 763 RVRSPISVTLK 773
Db 762 NVTSPLVVTILQ 772

```

GN MLN21.2/AT3G14240.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmid A.D., Lee J.M., Quach H.B., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene MLN21.2/AT3G14240 (GI:9279572).";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF360129; AAK25839.1; -.
DR HSP; O99405; 1MPT.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS50840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR KW Hypochemical protein.
FT VARIANT 5 S -> F.
FT VARIANT 5
SQ SEQUENCE 775 AA; 82524 MW; 52FA2C92B1C652C8 CRC64;

Query Match 42.3%; Score 1698; DB 10; Length 775;
Best Local Similarity 45.0%; Pred. No. 4.1e-107;
Matches 356; Conservative 134; Mismatches 257; Indels 44; Gaps 15;

QY 6 FFLCIIIFLLPCSSSSSEILQKQ---TYIVQLHPNSETAKTFASKFDWHLSTFQAVLGVEE 62
DB 3 FFFYFFELLTSSPSSASSNSLTIVVVDHEAKPS-IPFTHFWYTSLSASL- ---- 56

QY 63 EEEPSRLLYSGSAIEGFAAQLTSEAEILRYSPVAVRDPDHVLQVQTTYSYKFLGL 122
DB 57 --SSPPS-IHTYNTVHFPSARLTSDASQLLDHPHVISVPEQVRHLHTRSPEFLGL 113

QY 123 DGFNGSWGSKSPFGQGTIGVLDTGWPSPSFDGTGMPSPRKWKGIQCEGESFSSSS 182
DB 114 RSTDKAGLLEESDFGSDLVIGVDTGWMPSPSFDGTGMPSPRKWKGIQCEGESFSSSS 173

QY 183 CNKLCIGARFIRGHRVANSPESSPNPREYISARDSTGHTHTASTVGGSSVMANVLG 242
DB 174 CNKLVGARFCCGYEATNG---KMTETFEFRSDSDGHTHTASISGRVYFPASTILG 230

QY 243 NGAGVARGMAPGAHIAVYKVCWFGCYSSDILAADVAIQDKVDVLSLSLGGFPPIPLYDD 302
DB 231 YAHGVAAGMAPKARLAAYKVCWNSGCYSDILAADVAIDTAVADGVVLSVGGVVPYVLD 290

QY 303 TTAIGTFRAMEGISVICAGNNGPIESSVANTAPWVSTIGAGTLDRRPAVVRLANGKL 362
DB 291 AIAIGAFGALDRGIFVSASAGNGPGALITVNVAPWMTVAGTIDRDPFANVVLNGRM 350

QY 363 LYGESLYPGKIKNACREVEVIY---VTGCD-KGSEFFCLRGSLPREIEGKVIQDRGVN 418
DB 351 ISGVSVYGGPGL-DPGRMVPLTVYGGSLGSDGYSLSLCEGLSDPLNVKGLVLCDRGIN 409

QY 419 GRSEKGAIVEAGVAMILANTINOEDSIDVHLLPATLIGVTSVLKAVYNAVTK-- 476
DB 410 SRATKGEIVKNGGLGMIANGVDFDEGLVADCHVLPATSVGASGGDEIRVYSSSKSR 469

QY 477 ----PKARIIPGTVIGRSAPVAVAFSARGPSLANPSILKPDMAIPGVNIIAANPQNLG 532
DB 470 SSXHPATIVFKGTRLGIRAPVAVAFSARGPSLANPSILKPDMAIPGVNIIAANPDRIG 529

QY 533 PTGLPVDSSRVNTVMSGSCBHYSGITALLTRSAYPNWSPAAIKSAIMTTADLYDROG 592
DB 530 PSGVTSNRRNTEFNILSGTSMACPHVSGLAALLKAAHPDMSPAAIRSALITTAITYVDNSG 589

```

```

QY 593 KAIKD---GNKPAGVFAIGAGHVNPQKAINPGLVYNIQPDVITYLCTLGTFRSDILAIT 649
DB 590 EPMWDESTGN-TSSVWDYSGVHPFKANDPGLVVDITSVDYINFLCNSNYRTNIVITIT 648
QY 650 HKNVSCNGILRNKPGPSLNPYSIAVIFKRGKTEMIT---RRVTNVGSNSYSYVNVKAP 706
DB 649 RRQADCDGARRAGHVGNLNPFPSVVFQQYSGSKSTHFIRTVTNVGDSDSYVEIKIRPP 708

QY 707 EGIKTVNPKRLVFKHVDCTLSYRVWFLVKKNRGKVASFA----QGQLTWVNSHNLMO 762
DB 709 RGTIVTVEBEKUSFRRVGQKLS----FVVRVKTTEVLKSPGATNVETGHVWSGDG---KR 761
QY 763 RVASPISTVLK 773
DB 762 NVTSPVLVTLQ 772

RESULT 5
QBLGAG PRELIMINARY; PRT; 775 AA.
AC QBLGAG;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
DE Subtilisin-like serine protease.
OS Arabidopsis thaliana (Mouse-ear cress)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RA Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RL annotation.";
RN Genome Biol. 0:0-0(2002).
[2]
SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY084387; AAM60964.1; -.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS50840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Protease.
SQ SEQUENCE 775 AA; 82456 MW; 28E8753844348303 CRC64;

```

```

Query Match 41.9%; Score 1684; DB 10; Length 775;
Best Local Similarity 44.8%; Pred. No. 3.7e-106;
Matches 354; Conservative 133; Mismatches 260; Indels 44; Gaps 15;

QY 6 FFLCIIIFLLPCSSSSSEILQKQ---TYIVQLHPNSETAKTFASKFDWHLSTFQAVLGVEE 62
DB 3 FFFYFFELLTSSPSSASSNSLTIVVVDHEAKPS-IPFTHFWYTSLSASL- ---- 56

QY 63 EEEPSRLLYSGSAIEGFAAQLTSEAEILRYSPVAVRDPDHVLQVQTTYSYKFLGL 122
DB 57 --SSPPS-IHTYNTVHFPSARLTSDASQLLDHPHVISVPEQVRHLHTRSPEFLGL 113

QY 123 DGFNGSWGSKSPFGQGTIGVLDTGWPSPSFDGTGMPSPRKWKGIQCEGESFSSSS 182
DB 114 RSTDKAGLLEESDFGSDLVIGVDTGWMPSPSFDGTGMPSPRKWKGIQCEGESFSSSS 173

QY 183 CNKLCIGARFIRGHRVANSPESSPNPREYISARDSTGHTHTASTVGGSSVMANVLG 242
DB 174 CNKLVGARFCCGYEATNG---KMTETFEFRSDSDGHTHTASISGRVYFPASTILG 230

```


174	DB	CHNKLVGARFPCGGYEATNG---KWNETTEPRSPRSDSGHGHTWTSASIGRYYVFPASTJUG	230
243	QY	NGAGVARGMAPGAHIAVYKVCWFNGCXSSDILAAIDVAIQCDVDVLSLSLGGPFIPLYDD	302
231	DB	YAHGVAAGWAPKARILAAVKVCWNSGCYDSDILAFTDAVADGVDVLSLSVGGVVVPYLD	290
303	QY	TTAIGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTLDRRPPAVTVRLANGKL	362
291	DB	ATAIGAFAIDRGIFVFSASAGNGGFGALTVTNVPAPWMTTVGAGTIDRDPPANVKLGNGKM	350
363	QY	LYGESLYPEKGIKNAGREVEVIY---VTGGD-KGSECFCLRGSLPREBIEGRKMVICDRGVN	418
351	DB	IGSVSVYGGPGL-DGRMYPLVYCGSLGGDGYSSJCLEGSLDPNLVTGKIVLCDRGIN	409
419	QY	GRSEKGEAVKEAGGVAMILANTEINQEBSDIDVHLLPATLIGYTESVLLKAYVNAFVK--	476
410	DB	SRATKGEIVRKNGGLGMIANGVFDEGLVADCHVLPAFATSVGASGDEIRRYVSESSKSR	469
477	QY	---PKARIIFGGTIVGRSRAPEVAQFSARGPSLANPSILKPDWIAFGUNIIAAMPONIG	532
470	DB	SKKHPTATTIVFGTSLGIRPAPVASFARGNPETPEILKPDVIAPLGUNIILAAWPDRI	529
533	QY	PGLPYDGRVRNFTVMSGTSCPHVSGITALIRSAYPNWSPPAAIKSALMTADLYDROG	592
530	DB	PSGVTSIDNRTEFNILSGTSMACHVSGSLAALLKAAHPDWSPPAIRSALMTATYATVDNSG	589
593	QY	KAIKD---GNKPAGVAFAGAGHNPQKAINPGLVYNTQPVDIYTYCTIGTFRSDILAIT	649
590	DB	EPWMDSTGN-TSSVTDYSGSHVHPTRAMDPLGYDITSYDYNFLCNSNYGTNIVTIT	648
650	QY	HXNVSGNGLRKNPGFSLNYPISLAVIEPKRGKTTWMIT---RRVTNVSNSIYSVNVKAP	706
649	DB	RQACDGDGARRAGHVGUNLNPFSVVFQOYGESKXSTHFTRTVTNVDGSDSVYEIKIRPP	708
707	QY	EGIKVIVNPKRLVFXHVDQTLISYRVWFVLKKNRGGKVASFA---CGQ.TWVNSHNLMQ	762
709	DB	RGITVTVPEKLSFRVQGLS---FVVRVKTTTEVKLSPGATNVTGHWMSDG---KR	761
763	QY	RVRSFISVTLK	773
762	DB	NVTSPVLVTLQ	772

[illegible]

[illegible]

Serine protease-like protein (Putative subtilisin serine protease).
 GN AF551750.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopses.
 RX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=98290546; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.,
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned P1 and TAC clones."
 RL DNA Res. 5:41-54(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya M., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RT "Arabidopsis Full Length cDNA Clones."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB010074; BAB1244.1; -;
 DR EMBL; AY080831; AAL87307.1; -;
 DR HSP; Q45670; IDBI.
 DR MEROPS; S08.092; -;
 DR InterPro; IPR000183; Decarboxylase2.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_s8.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_s8; 2.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00879; ODR_DC_2_2; 1.
 DR PROSITE; PS00840; PA; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Protease.
 SQ SEQUENCE 780 AA; 84950 MW; BE22D41550C8BCFC CRC64;
 Query Match 40.9%; Score 1642.5; DB 10; Length 780;
 Best Local Similarity 43.6%; Pred. No. 2.5e-103;
 Matches 342; Conservative 128; Mismatches 281; Indels 33; Gaps 11;
 QY 4 KPFFLCI---IFLLFCSSSEILQKQTVIVQLHPNSETAKTFASKDWHLSFLOEAVLG 59
 DB 9 KPFLFILNLIPLQAEITVQISTKTKYVIHM-DKSMPLPYTNHLQWYSSKINSVTOH 67
 QY 60 VEEEEPEPRLLYSGSAIEGPAALQTESEAILRYSPEVAVDPDHLVQVOTVSYKF 119
 DB 68 KSOEEEGNNRIITYQTAFHGLAQLTOEEAEERLEEEDGVAVIPETRYELHTRSTFF 127
 QY 120 LGLDGFGNSGVWKSFRGGQTIIGVLDGTGWPEPSFDITGMPSTPRKWKGCQCEGSEFS 179
 DB 128 LGLERQESRVAERVTDDVVGVLDGTGWPESESNFTGMSVPFATWRGACETGKRF 187
 QY 180 SSSCNKRLIGARFIFGRHVANSPEPSNMPREYISARSTGCHGTHASTVGGSSVSMAN 239
 DB 188 KNCNKRKIVGARFYRYGEAATGKIDE---ELEYKSPDRDRGDGHGTHAATVAGSPVKGAN 244
 QY 240 VLNGAGVARGMAPGAHIAVYKVCWPGNGCVSSDILAAIDVAIQDQVDVLSLGLGFPPL 299
 DB 245 LFGFAGTAGMAQKARVAYKVCWVGCGCFSSDILSAVDQAVDGVQVLSISLGGGVSTY 304
 QY 300 YDTIATGTFRAMEGIVICRAGNNGPTESSVANTAPVWSTIGAGTLDRRPPAVVRLAN 359
 DB 305 SRDLSIAITFGAMGVFFVSCSAGNGGPDPISLITVNSPWITTVGASTMDRDPATVKIGT 364

360 GKLYGSESLYPGKIKNAGREVVIYV---TGGDKGSEFCLRGSLPREIRGRKMWICDRG 416
 DB 365 MRTEKGVSLYKGRTVLPKNQKQYPLVYLGRLNASSPDP*SFCLDGDALDRHVAGKIVICDRG 424
 QY 417 VNGSRKEAEVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGYTESVLLKAYVNAIVK 476
 DB 425 VTPRVQGVVVRAGGIGMVLITNTATNGBELVADSHMLPAVAVGEGKGLIKQYAMTSKK 484
 QY 477 PKARIIFGCTVIGRSRAPEVAQFSARGPSLANPILKPDMLAPGVNIIAAMPONLGTGL 536
 DB 485 ATASLEILGTRIGIKPSFVVAFAFSRGNFSLLEILAPDLLAPGVNIIAAMPONLGTGL 544
 QY 537 PYDSRRNVFTVMSGTSMSCHVSGITAIIRSAVNPWSPAAIKSALMTTADLYDQGGKAIK 596
 DB 545 SSDPRRVKFNLSGTSMSCHVSGVAALIKSRHDPWSPAAIKSALMTTAYVHDNMFKPLT 604
 QY 597 D--GNKPAGVPAIGAGHVNPKAINPGLVNIQPDYITVLTGLTFRSILAIT-HQNV 653
 DB 605 DASGAAPSSPYDHGAGHIDPLRATDPLGVYDIGQEYFEFLCTQDLSPLSKVFTKHSNR 664
 QY 654 SCNGILRNKPGFSLNYPFSAIVIPKRGKTTMIT--RRVTNVGSPNSIYSVNVKAPEGIKV 711
 DB 665 TCKITLAKNPG-NLNYFAISALFPENTHVKAMTLRRVTNVGPHISSYKVSVPFRGASV 723
 QY 712 IVPNKRLLVFKHVDQTLISY---RVWFVLYKKNRGGKVASFAQGQLTVWNSHNLMOVRSP 767
 DB 724 TVQPKTLNFTSKHOKLSYTVTFTRFRMKRPEFG-----LVWKSIT--THKVRSP 771
 QY 768 ISVT 771
 DB 772 VIIT 775

RESULT 12
 Q39007
 ID Q39007 PRELIMINARY; PRT; 746 AA.
 AC Q39007;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Subtilisin-like protease (Fragment).
 GN ARA12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopses.
 CX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=95375542; PubMed=7647567;
 RA Ribeiro A., Akkermans A.D.L., Van Kammen A., Bieseling T.,
 RA Pawlowski K.,
 RT "A nodule-specific gene encoding a subtilisin-like protease is
 RT expressed in early stages of actinorhizal nodule development."
 RL Plant Cell 7:785-794(1995).
 RN [2]
 RP SEQUENCE FROM OF 481-613 N.A.
 RC STRAIN=cv. Columbia; TISSUE=Dry seed;
 RA Raynal M., Grellet P., Laudie M., Meyer V., Cooke R., Delseeny M.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X85974; CAA59963.1; -;
 DR EMBL; Z47391; CAA87460.1; -;
 DR HSP; Q45670; IDBI.
 DR InterPro; IPR00183; Decarboxylase2.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_s8.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_s8; 3.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00879; ODR_DC_2_2; 1.
 DR PROSITE; PS00840; PA; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.

[illegible]


```

RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shimizu P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene At4g34980 (GI:15236234).";
RN Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL [6]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shimizu P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RT "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL020223; CAA17763.1; -
DR EMBL; AL161586; CAB80215.1; -
DR EMBL; AY074375; AAL67071.1; -
DR EMBL; AY096357; AAM19998.1; -
DR HSSP; O99405; 1MPT.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 764 AA; 81046 MW; EBB38272F6F33F50 CRC64;

Query Match 39.5%; Score 1589; DB 10; Length 764;
Best Local Similarity 44.0%; Pred. No. 1.1e-99;
Matches 342; Conservative 127; Mismatches 268; Indels 40; Gaps 16;

QY 7 FLCIIIFLLPCSSSEILQQTYYIVQLHPNSETAKTFASKFDWHLFLQEAVALGVEEBEE 66
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
11 FLSEPTISEAASQA---AKTFIFRIDGGS-WPSIFPTHYHWY-----STFEAEE- 55

QY 67 PSSRLLYSGSAIEGFAAQLTSEAEILAYSPEWAVRDPHVLOVQTTYSYKFLGLDGF 126
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
56 --SRIVHVHTVPHGFSVAVTPEADNLNHPAVLAVFEDRRRLHTRSPQFLGLQ--N 111

QY 127 NSGWKSKRFGQGTIIIGVLDTGWPSPSPGDDTGMPISIPRKWKIGICEGESFSSSCNRK 186
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
112 QKGLWSESDYGVIIIGVPTGTWPERRSFSDLNLOPIPKRWGVCESARSPNCNRK 171

QY 187 LIGARFFIRHVRANSPSPNMPREYISARDSTGHGTHASTVGSSVSMANVLGNAG 246
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 IIGARFFAKQQA--VIGGINKTVEFLSPRADGCHGTHSTSTAAGRHAFAKASMGYASG 229

QY 247 VARGMAPGAHAYKVCW-FNGCYSSDILAAIDVAIQDVVLSLSLG---GFPPIYDD 302
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
230 VARGVAPKARIAAYKVCWDSGLSDILAAFDAAVAVRDGVDVISISIGGGDITSPYILD 289

QY 303 TIAIGTFRAMEGIVTCAAGNNGPIESSVANTAPWVSTIGAGTLDLDRPPAVVRLANGKL 362
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
290 PIAIGVGAASKGIFVSSSAGNEGPNMGSMVTNLAPWTTVVGASTIDRNPDAIIGDHR 349

QY 363 LYGESLYPGKGIKNAGREVEIYV-TGGDKGSEFCLRGSLPREIRKGMVICDRGVNGRS 421
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
350 LRGVSLYAGVPLN--GRMFPVYVPGKSGWSSASLCHMENTLDPKVRGKIVICDRGSSPRV 407

QY 422 EKGEAVKEAGGVAMILANTEINQEDSIDVHLLPATLIGYTESVLLKAYVNAVTVKPKARI 481

```

Search completed: February 2, 2004, 09:11:26
Job time : 45 secs

```

Db 408 AKGLVVKAGGVGMILANGASNGEGLVGDHLLIPACAVGNSNEGDRIKAYASHPNPIASI 467
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
482 IFGGTVIGRSRAPEVAOFSARGPSLANPSILKPDMIAPGWNIIAAWPNQGLTGLPYDSR 541
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
468 DFRGTVIGIKPADVIASFSGRPGNGLSPILLKPDILAPGVNIIAAWTDVAGTGLPSDR 527
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
542 RVNFTVMSGTSMSCPHVSGITALIRSAFENWSPAIAKSAALMTTADLYDSQKAIKDN-- 599
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
528 KTEFNILSGTSMACPHVSGAAALLKSAHPDWSPAVIRSAMMTTTLNLDVNSRSLIDESTG 587
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
600 KPAGVFAIGAGHVNPOKAINPGLVYNIQPDYITLYCTLGFTSRSDILAIHKNVSCNGIL 659
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
588 KSATPDYDGGSHLNLGRANPGLVYDITNDYITFLCSIGYGPKTIVITRTPVRCPTTR 647
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
660 RKNPGFSLNYPSTAVIF--KRGKTEMITRRVTNVGSPNSIYSVNVKAPGEGIKVIVNPK 716
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
648 KPSFG-NLNTPSITAVPTNRRGLVSKTVIRTAINVGOAEVYRARIESFRGVTVTVKPP 706
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
717 RLVEFKHVDQTLSYRVNMFVLKKKCN--RGGKVASFAQQLTWNASHNLMQVRSPISVT 771
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
707 RLVEFSAVKERSYAVITVTNTRNVVLGETGAVF--GSVTFDGGKHV--VRSPIVVT 759

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 3, 2004, 01:46:31 ; Search time 2735 Seconds
(without alignments)
6887.009 Million cell updates/sec

Title: US-09-806-767-2
Perfect score: 4018
Sequence: 1 MEKPPFLCIIIFLLFCSSS.....NSHNLQVRSPISVTLKTN 775

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DSV=xlh
-Q=/cgn2/1/USPTO.spool/US09806767/runat_02022004_085958_17911/app_query.fasta_1.967
-DB=EST -OPMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCAL=CN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPM=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09806767@cgn_1_12132@runat_02022004_085958_17911 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1282	31.9	832	28	BH480160
2	1092	27.2	696	28	BH660501
3	1079.5	26.9	1572	11	AY107161
C 4	1073	26.7	676	28	BH998381
C 5	1054	26.2	670	28	BH464077
C 6	1012	25.2	635	28	BH59290
C 7	979	24.4	640	28	BH442444
C 8	961.5	23.9	789	29	CC340521
9	923	23.0	627	28	BH597572
10	887.5	22.1	1795	11	AY108210
C 11	863	21.5	803	28	BH601308
12	820	20.4	750	28	AZ046831
13	775	19.3	767	28	BH793726
14	764	19.0	568	10	BF052773
C 15	749	18.6	527	28	BH714875
C 16	744	18.5	1006	11	AY105444
17	728.5	18.1	713	28	BH720365
18	715.5	17.8	719	13	BQ870650
19	700.5	17.4	844	14	CD434609
20	698	17.4	848	14	CD575777
21	696.5	17.3	849	14	CB642755
22	683.5	17.0	800	14	CB892659
23	680	16.9	859	14	CD573880
24	678	16.9	740	13	BQ867188
25	677	16.8	863	10	BF278122
26	676	16.8	873	10	BE216378
27	673.5	16.8	818	14	CD437637
28	669.5	16.7	697	29	BZ403171
C 29	665.5	16.6	710	13	BQ507537
30	662.5	16.5	768	29	BZ536866
31	661	16.5	709	13	BQ866922
C 32	656	16.3	853	28	BH649993
33	653	16.3	744	14	CB982404
C 34	651.5	16.2	878	29	BZ656313
35	646	16.1	738	13	BQ996378
36	636	15.8	770	10	BG598379
37	634	15.8	771	13	BU001251
38	632	15.7	771	13	BQ992191
39	630.5	15.7	926	10	BG445439
40	627.5	15.6	762	28	BH649984
41	627.5	15.6	923	29	BZ786320
C 42	625.5	15.6	766	28	BH725669
43	625.5	15.6	818	12	B1405859
C 44	624	15.5	864	28	BH648619
45	621.5	15.5	811	12	BG887274

ALIGNMENTS

RESULT 1
BH480160/c BH480160 832 bp DNA linear GSS 13-DEC-2001
LOCUS BOHME48TF BOHM Brassica oleracea genomic clone BOHME48, genomic
DEFINITION survey sequence.
ACCESSION BH480160
VERSION BH480160.1 GI:17688264
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.

```

REFERENCE      1 (bases 1 to 832)
AUTHORS        Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE          Whole genome shotgun sequencing of Brassica oleracea
JOURNAL        Unpublished
COMMENT        Other_GSSs: BOHME48TR
                Contact: Chris Town
                TIGR
                9712 Medical Center Drive, Rockville, MD 20850, USA.
                Tel: 301-838-3523
                Fax: 301-838-0208
                Email: cdtown@tigr.org
                DNA is from a doubled haploid provided by Tom Osborn.
                Seq primer: TF
                Class: sheared ends.

FEATURES             Location/Qualifiers
     source           1..832
     /organism="Brassica oleracea"
     /mol_type="genomic DNA"
     /strain="TO1000DH3"
     /db_xref="taxon:3712"
     /clone="BOHME48"
     /note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
     genomic DNA inserted into pHOSt using BstXI linkers"
BASE COUNT      177 a 209 c 247 g 199 t
ORIGIN
1
Alignment Scores:
Pred. No.:      1,21e-123      Length:      832
Score:          1282.00        Matches:     244
Percent Similarity: 93.84%      Conservative: 15
Best Local Similarity: 88.41%    Mismatches:  16
Query Match:     31.91%         Indels:       1
DB:              28            Gaps:          0

US-09-806-767-2 (1-775) x BH480160 (1-832)

QY      83 AlaAlaGlnLeuThrGluSerGluAlaGluLeuLeuArgTyrSerProGluValValAla 102
Db      832 TCAGCTCAGTTACCGAATCAGAACCCAAACACTCAAGACTTTACCTGCAAGTTGTTGCA 773
QY      103 ValArgProAspHisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeu 122
Db      772 GTAAGACCTGACCAATGTTCTCAAGTCCAAACACTTACTCTTACAAAGTTCTTAGGACTC 713
QY      123 AspGlyPheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIle 142
Db      712 AACGGTCCAGGACCTCCAGCGTATGTTCTAAATCCGGTCCGGCCAAAGGCACGATCATC 653
QY      143 GlyValLeuAspThrGlyValTrpProGluSerProSerPheAspAspThrGlyMetPro 162
Db      652 GCGGTTCTCGACACAGGAGTCTGGCCCGAAAGCCCTAGCTTCGACGACACAGGGATGCT 593
QY      163 SerIleProArgLysTrpLysGlyIleCysGlnGlyGlySerPheSerSerSerSer 182
Db      592 TCAGTCCCAAGCAAAATGGAAGAGTGTTGCCAAGAGGAGAGAGTTCATCTTCCTCAAAC 533
QY      183 CysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSer 202
Db      532 TGTAAAGAAACTAATCGCGCTAGATTCTTCATCAGAGGCCACCGCGTAGCCAACTCT 473
QY      203 ProGluGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHis 222
Db      472 CCTTTGAGACTCAACAAATGCTCGCAATACATCGGCACGAGACTCCACGGGGCAC 413
QY      223 GlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGly 242
Db      412 GGGACCCACACCCCTCAACTGCTGGCGGCTCCTCTGTTTCAGCCGCGAGCGTTCTCGGT 353
QY      243 AsnGlyValAlaGlyValArgGlyMetAlaProGlyValaHisIleAlaValTyrLysVal 262
Db      352 AACGGAGCCCGGGTAGCCCGCGGATGGCCCGCAGGAGCTCAGCTGGCGGTTTTACAAAGTC 293

```

```

QY      263 CysTrpPheAsnGlyCysTyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGln 282
Db      292 TGTGTTTCAACGGCTGTTACAGCTCAGACATTCGGCGGCATAGACGATACGATACAA 233
QY      283 AspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAsp 302
Db      232 GACAAAGTCGAGTCTATCCCTCTCCCTAGCGGGTTCCCTATCCCTTTGATGATGAC 173
QY      303 ThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAla 322
Db      172 ACTATCGCCATGGTACCTTCGGCGCCACGGAACATCGAATCTCGTTGTCGGCGGT 113
QY      323 GlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIle 342
Db      112 GGCACAAACGGCGGCTCGCTAGCTCGTGGCGGAACACAGCTCCTTGGGTCTCAACCATC 53
QY      343 GlyAlaGlyThrLeuAspArgPhePro-AlaValValArgLeu 357
Db      52 GGAGCAGGAACGCTCGATAGAAATTCGGCGGGGTGTTAGGTTA 7

RESULT 2
BH60501
LOCUS          BH60501
DEFINITION    BOHWS71TF BO 2.3 KB Brassica oleracea genomic clone BOHWS71,
                genomic survey sequence.
ACCESSION     BH60501
VERSION       BH60501.1 GI:18719169
KEYWORDS      GSS.
SOURCE        Brassica oleracea
ORGANISM      Brassica oleracea
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
                ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE     1 (bases 1 to 696)
AUTHORS       Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE         Whole genome shotgun sequencing of Brassica oleracea
JOURNAL       Unpublished
COMMENT       TIGR
                Contact: Chris Town
                9712 Medical Center Drive, Rockville, MD 20850, USA.
                Tel: 301-838-3523
                Fax: 301-838-0208
                Email: cdtown@tigr.org
                DNA is from a doubled haploid provided by Tom Osborn.
                Seq primer: TF
                Class: sheared ends.
                Location/Qualifiers
                 source
                 1..696
                 /organism="Brassica oleracea"
                 /mol_type="genomic DNA"
                 /strain="TO1000DH3"
                 /db_xref="taxon:3712"
                 /clone="BOHWS71"
                 /clone_lib="BO 2.3 KB"
                 /notes="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
                 genomic DNA inserted into pHOSt using BstXI linkers"
BASE COUNT    163 a 186 c 220 g 147 t
ORIGIN
1
Alignment Scores:
Pred. No.:      8.09e-104      Length:      696
Score:          1092.00        Matches:     206
Percent Similarity: 95.22%      Conservative: 13
Best Local Similarity: 89.57%    Mismatches:  11
Query Match:     27.18%         Indels:       0
DB:              28            Gaps:          0

US-09-806-767-2 (1-775) x BH60501 (1-696)

QY      229 ThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAla 248
Db      6 ACTGTCGCGGCTCCCTCTGTTTCGACGCGAGGCTTCTCGGTAAACGAGCGGGTAGCC 65

```

```

QY 249 AtgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys 268
Db 66 CGCGGATGCGGCAGAGCTACGTCGGGGTTTACAAAGTCTGCTGTTCAACGGCTGT 125
QY 269 TyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeu 288
Db 126 TACAGCTCAGACATTCTGGCGGGATAGAGTAGCGATACAGACAAAGTCAGCTTCTA 185
QY 289 SerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThr 308
Db 186 TCGCTCTCCCTAGCGGTTCCCTATCCCTTTGTATGATGACATATGCCATTCGGTACC 245
QY 309 PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle 328
Db 246 TTCGCGGCCACGGAACATCTCGTTGTCTGCGGGGTGGCAACACGCGCGCTC 305
QY 329 GluSerSerValAlaAsnThrAlaProTyrValSerThrIleGlyAlaGlyThrLeuAsp 348
Db 306 GCTAGTCTGGTGGCAACACAGCTCTTGGTCTCAACCATCGGACGAGCAACGCTCGAT 365
QY 349 ArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeu 368
Db 366 AGAAATTCCTCGGGGTTGTAGTTAGCAACGGGAGCTACTCTACGAGAGTCGTG 425
QY 369 TyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGly 388
Db 426 TACCAGGGAAGGATAGAGAGCTGAGAGAGTTGGAGATTCTACCTGCGGGA 485
QY 389 GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGly 408
Db 486 GGAGACAAAGGAGGAGATTTCTGAGAGGCTCGCTTCCGAGAGAGTGTCCAAAGC 545
QY 409 LysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLys 428
Db 546 AAGATGGTATATGCCAGCTGGAGTCACGGGAGATCAGAGAGGACAGCGGTTAAA 605
QY 429 GluAlaGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSer 448
Db 606 GAAGCTGGCGGCTCGGATGATCTTGGCTTAACACCGAGATTAAACCAAGAGGAGATTCT 665
QY 449 IleAspValHisLeuLeuProAlaThrLeu 458
Db 666 GTAGACGTTTCATTGTATACCAGCTACGTTG 695

RESULT 3
AY107161 1572 bp mRNA linear HTC 16-OCT-2002
LOCUS Zea mays PC0101731 mRNA sequence.
DEFINITION Zea mays PC0101731 mRNA sequence.
ACCESSION AY107161
VERSION AY107161.1 GI:21210239
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1572)
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 1572)
Coe, E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:

```

```

www.zmdb.iastate.edu.
FEATURES
Location/Qualifiers
1..1572
/mol_type="mRNA"
/db_xref="MaizeDB:635852"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"
/notes="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"
BASE COUNT 297 a 542 c 465 g 268 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 5.66e-102 Length: 1572
Score: 1079.50 Matches: 230
Percent Similarity: 61.72% Conservative: 86
Best Local Similarity: 44.92% Mismatches: 177
Query Match: 26.87% Indels: 19
DB: 11 Gaps: 10

```

US-09-806-767-2 (1-775) x AY107161 (1-1572)

```

QY 268 CysTyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspVal 287
Db 40 TGTCTTACGTCGACATCTCTGGCGGCGTCGACCGCGCTGGCGGACGCGCTGACGTG 99
QY 288 LeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGly 307
Db 100 CTCTCCATCTCGCTCGCGGCGGCTGCTCCCTCTACTTCGCGACAGCTGGCCATCGCG 159
QY 308 ThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyPro 327
Db 160 TCGTTCGTGTCATGCGATGGCGCTGCTCGCTGCTGCTGGCGGCGCAACGCGGCGCG 219
QY 328 IleGluSerSerValAlaAsnThrAlaProTyrValSerThrIleGlyAlaGlyThrLeu 347
Db 220 GACCCCAATAAGCTTCACCAACCTGTCTGCGCTGGATACCAACGCTGGCGGCGGACCATG 279
QY 348 AspArgArgPheProAlaValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSer 367
Db 280 GACCGGAGCTTCCCGGCGACGCTGCGCAATGGCGCAACCTTCACCGGGGTTTCG 339
QY 368 LeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThr 387
Db 340 CTCTACAAGCGCGCGAGGTCTTCTGCTCAAGGAGCAGTACCCGTTAGTCTACATGGCG 399
QY 388 GlyGlyAspLys-----GlySerGluPheCysLeuArgGlySerLeuProArgGlu 404
Db 400 GGCAACTCGAGCATCCCTGACCCCGAGTCCCTGCTGCTGGAGGGGACACTCCAGCCGCCAC 459
QY 405 GluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGly 424
Db 460 GAGTCTCGCTGAAAGATTGTGATCTGCGACGCGGCGCATTTAGTCTCGGGTGCAGAGGGT 519
QY 425 GluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGln 444
Db 520 CAGTTTGTCAAGAACCGCGGTGCGCGGCGATGATCTCGCCCAACACCGCGCAACCGCG 579
QY 445 GluGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSer 464
Db 580 GAGAGGTTGTCGCGGACACCCACTCTCTCCAGCAGTGGCCGCTGGCGAGTCTGAAGCG 639
QY 465 ValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGly 484
Db 640 ATTCCCGCCAAAGAGTACAGCAAAACCGCCCGCAACCAACGCGGCGAGCTCAGCTTCGAC 699
QY 485 GlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyPro 504

```

```

700 GGGACGAAGTCGGGATCCGCCATCGCCAGTCGTAGCGCGTCTTCGTCGCCGGGACCA 759
505 SerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsnIle 524
760 AATCTCTGACCTCGAGATCTCTCAAGCGGAGTTCATCGCGCGCGGTGAACATCTG 819
525 AlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsn 544
820 CGGCATGGAGCGCGACCGCCGTCGAGCTTGTCCAGCAGCGCGCGGGTGGC 879
545 PheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeu 564
880 TTCAACATCTGTCGGGAGCTCATGTCGTGCCGACGTCGCGCGCTGCGCTGCGCTG 939
565 IleArgSerLysProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThr 584
940 ATCAAGCCGACCGGACCGGACCGCGCGGAGATCAAGTCGCGCTGATGACCCACC 999
585 AlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAla 602
1000 GGTACGTCACACACACACACGTACCGTGGCTCGTAGAGAGCGGCGCCACCGGAGCGTCC 1059
603 GlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeu 622
1060 ACGCGTTCGATCAGCGAGCTGGGCACATACACCGCTGCTGTCCTCAACCTGCGCTG 1119
623 ValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArg 642
1120 GTCTACGACATCGCCAGGACGACTACCTGGAGTTCCTCTGCTGGAGAACCTGACGCG 1179
643 SerAspIleLeuAlaIleThrHisLysAsnValSer-----CysAsnGlyIleLeuArg 660
1180 TTGACGCTCAGGCTCTTCCACC---AAGACTCGAGCAAGACATGCAAGCACCTTC--- 1233
661 LysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
1234 AGTCGCGCGGT---GACTTGATTATCCGGCCATCTCCGCGTCTTCGCGGAGCAGCCA 1290
681 ThrThrGluMetIleThrArgArg---ValThrAsnValGlySerProAsnSerIleTyr 699
1291 TCTGCTCGCTGACGCGTTCGACCGCTGACGAGCGTTCGCGCGCGCGCTTCGACTTAC 1350
700 SerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuVal 719
1351 CATGTCGAAGTTACAGATTCAAGGCGCAGACATTCGTCGAACCAAGACACCTTGCAC 1410
720 PheLysHisValAspGlnThrLeuSerTyrArgValTyrPheValLeuLysLysAsn 739
1411 TTCACAAGTTCGACACCAAGAGTAACTACAGGTGACGATGACACCAAGCGTGGCCAG 1470
740 ArgGlyLysValAlaSerPheAlaGlnGlyLysLeuThrTrpValAsnSerHisAsn 759
1471 -----AAGACACCGGAGTTC-----GGAGCGCTGCTTTGAGCAGCGCGTCCAC 1515
760 LeuMetGlnArgValArgSerProIleSerValThr 771
1516 ATC-----GTCCGGAGCCCGCTTGTCTCCACA 1542

RESULT 4
LOCUS BH998381/c
DEFINITION oeh58ell.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
ACCESSION BH998381
VERSION BH998381.1
KEYWORDS GI:23544689
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 676)

```

```

AUTHORS Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T., Nash
W., Rabinowicz, P.D. and Willson, R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Richard K. Willson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Plate: oeh58 row: e column: 11
Seq primer: -21UPpOT forward
Class: shotgun
High quality sequence start: 14
High quality sequence stop: 551.
FEATURES
Location/Qualifiers
1..676
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."
BASE COUNT 138 a 177 c 208 g 153 t
ORIGIN
Alignment Scores:
Pred. No.: 7.64e-102 Length: 676
Score: 1073.00 Matches: 200
Percent Similarity: 94.20% Conservative: 11
Best Local Similarity: 89.29% Mismatches: 13
Query Match: 26.70% Indels: 0
DB: 28 Gaps: 0
US-09-806-767-2 (1-775) x BH998381 (1-676)
QY 118 LysPheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysArgPheGly 137
DB 675 AAGTTCTTAGACTCAACGGTCCAGGACCTCCAGCGTATGGTCTAAATCCGCTCCGCG 616
QY 138 GlnGlyThrIleLeGlyValLeuAspThrGlyValTrpProGluSerProSerPheAsp 157
DB 615 CAAGCGCAGATCATCGCGGTTCGACACAGGAGTCTGGCCCAAGAGTGTTCGCAAGAGGAGAGC 556
QY 158 AspThrGlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGlyGluSer 177
DB 555 GACACAGGATGCTTCTCAGTCCCAAGCAATGGAAAGTGTTCGCAAGAGGAGAGC 496
QY 178 PheSerSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHis 197
DB 495 TTCATTCTCTCAAACTGTAAACAGAACTAATCGCGCTAGATTCTTCATCAGAGGCGAC 436
QY 198 ArgValAlaAsnSerProGluGluSerProAsnMetProArgGluTyrIleSerAlaArg 217
DB 435 CGCGTAGCCAACTCTCTTTGGACTACCAAAACATGCTCGCGAATACATATCGGCACGA 376
QY 218 AspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMet 237
DB 375 GACTCCACGGGCGACGGAGCCACACCGCTCAACTCTGCGCGCTCTCTGTTTCACG 316
QY 238 AlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIle 257
DB 315 CGGAGCGTCTCTCGTAAACGGAGCGCGGTAGCCCGCGGGATGGCCGAGAGCTCAGCTG 256
QY 258 AlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAspIleLeuAlaIle 277
DB 255 CGCGTTTACAAAGTCTCTGCTTCAACGGCTGTGTACAGCTCAGACATCTCTGGCGCGATA 196
QY 278 AspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyPheProIle 297

```

```

AUTHORS Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T., Nash
W., Rabinowicz, P.D. and Willson, R.K.
TITLE Whole genome shotgun reads from Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Richard K. Willson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Plate: oeh58 row: e column: 11
Seq primer: -21UPpOT forward
Class: shotgun
High quality sequence start: 14
High quality sequence stop: 551.
FEATURES
Location/Qualifiers
1..676
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."
BASE COUNT 138 a 177 c 208 g 153 t
ORIGIN

```

```

Db      195 GAGGTAGCATACAGACAAAGTCGACGTTCTATCCCTCTCCCTAGGCGGTTTCCTATC 136
Qy      298 ProLeuTyraSPThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSer 317
Db      135 CCTTGTATGATGACACTATGCCATTGGTACTTCCGCGCCACGACATGGAATCTCC 76
Qy      318 ValIleCyAlaAlaGlyAsnGlyProIleGluSerSerValAlaAsnThrAlaPro 337
Db      75 GTTGTCTGCGCGGTGGCAACACGCGCGCTCGCTAGCTCGGTGGCAACACAGCTCCT 16
Qy      338 TrpValSerThr 341
Db      15 TGGATTCCACC 4

RESULT 5
BH464077/c
LOCUS      670 bp      DNA      linear      GSS 13-DEC-2001
DEFINITION BOGEF01TF BOGE Brassica oleracea genomic clone BOGEF01, genomic
survey sequence.
ACCESSION      BH464077
VERSION      BH464077.1 GI:17659406
KEYWORDS      GSS
SOURCE      Brassica oleracea
ORGANISM      Brassica oleracea
REFERENCE      1 (bases 1 to 670)
AUTHORS      Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE      Whole genome shotgun sequencing of Brassica oleracea
JOURNAL      Unpublished
COMMENT      Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TP
Class: sheared ends.
FEATURES             Location/Qualifiers
     source            1..670
     /organism="Brassica oleracea"
     /mol_type="genomic DNA"
     /strain="TO1000DH3"
     /db_xref="taxon:3712"
     /clone="BOGEF01"
     /note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
     genomic DNA inserted into pHOSt using BstXI linkers"
BASE COUNT      137 a 188 c 186 g 159 t
ORIGIN
Alignment Scores:
Pred. No.:      7,44e-100      Length:      670
Score:      1054.00      Matches:      197
Percent Similarity:      94.14%      Conservative:      12
Best Local Similarity:      88.74%      Mismatches:      13
Query Match:      26.23%      Indels:      0
DB:      28      Gaps:      0

US-09-806-767-2 (1-775) x BH464077 (1-670)

Qy      158 AspThrGlyMetProSerIleProArgLysTrpLysGlyIleCyAlaGlyGluSer 177
Db      670 GACACAGGATGCTTTCATCCACCAATGAAAGGTGTTGCCAAGAGGAGAGAC 611
Qy      178 PheSerSerSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHis 197
Db      610 TTCACTCTTCAACTGTAACAAGAACTAATTCGCGGTAGATTCTTCATCAGAGGCCAC 551
Qy      198 ArgValAlaAsnSerProGluSerProAsnMetProArgGluTyrlleSerAlaArg 217

```

```

Db      550 CGGTAGCAACTCTCTTTGGACTCACCACAAATCATCCCTCGGAATACATATCGGCACGA 491
Qy      218 AspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMet 237
Db      490 GACTTCACGGGGCACGGGACCCACACCGCTCAACTGCTGGCGGCTCTCTGTTCGACG 431
Qy      238 AlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIle 257
Db      430 GCGAGCGTTCCTCGGTAAACGAGCGGGGTAGCCCGCGGGATGGCGCCAGGAGCTCACGTG 371
Qy      258 AlavalTyrlleValCysTrpPheAsnGlyCysTyrlleSerSerAspIleLeuAlaIle 277
Db      370 GCGTTTACAAAGTCTGCTGGTTCAACGGCTGTGTACAGTCAACATTTCTGGCGGCAATA 311
Qy      278 AspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIle 297
Db      310 GAGTAGCGATACAAACAAAGTCGAGCTTCATCTCTCCCTAGGCGGTTTCCCTATC 251
Qy      298 ProLeuTyraSPThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSer 317
Db      250 CCTTTGTATGATGACACTATCGCAATTGGTACCTTCGCGCCACCGAATGGAATCTCC 191
Qy      318 ValIleCyAlaAlaGlyAsnGlyProIleGluSerSerValAlaAsnThrAlaPro 337
Db      190 GTTGTCTGCGCGGTGGCAACACGCGCGCTCGCTAGCTCGGTGGCGAACACAGCTCCT 131
Qy      338 TrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValAlaArgLeu 357
Db      130 TGGGTCTCAACATCGAGCGAGACGCTCGATAGAAATTCGCGGGGTGTTAGTTA 71
Qy      358 AlaAsnGlyLysLeuLeuTyrlleGlyGluSerLeuTyrlleGlyLysGlyIleLysAsnAla 377
Db      70 GCCACGGGAAGCTACTCTACGGAGAGTCTGTACCCAGGAAAGGATAGAGAGCGGA 11
Qy      378 GlyArg 379
Db      10 GAGAGA 5

RESULT 6
BH59290/c
LOCUS      635 bp      DNA      linear      GSS 15-DEC-2001
DEFINITION BOGWR58TR BOGW Brassica oleracea genomic clone BOGWR58, genomic
survey sequence.
ACCESSION      BH59290
VERSION      BH59290.1 GI:17851736
KEYWORDS      GSS
SOURCE      Brassica oleracea
ORGANISM      Brassica oleracea
REFERENCE      1 (bases 1 to 635)
AUTHORS      Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE      Whole genome shotgun sequencing of Brassica oleracea
JOURNAL      Unpublished
COMMENT      Other_GSSs: BOGWR58TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES             Location/Qualifiers
     source            1..635
     /organism="Brassica oleracea"
     /mol_type="genomic DNA"
     /strain="TO1000DH3"
     /db_xref="taxon:3712"
     /clone="BOGWR58"
     /clone_lib="BOGW"

```

/note="Vector: pHS01; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHS01 using BstXI linkers"

BASE COUNT

127 a 198 c 162 g 148 t

ORIGIN

Alignment Scores:

Pred. No.: 1.75e-95 Length: 635
Score: 1012.00 Matches: 190
Percent Similarity: 94.7% Conservative: 10
Best Local Similarity: 90.0% Mismatches: 11
Query Match: 25.1% Indels: 0
DB: 28 Gaps: 0

US-09-806-767-2 (1-775) x BH599290 (1-635)

QY 213 TTTTCTGTTTTCAGCGGAGGCTTCTCGGTAAACGAGCGGGGTAGCCCGGGGATGGCG 514
Db TACATATCGCAGAGACTCCACGGCCACCGGACCCACCGCTCAACTGCTGGCGGC 574
QY 233 SerSerValSerMetAlaAsnValLeuGlyAanGlyAlaGlyValAlaArgGlyMetAla 252
Db TCTCTGTTTTCAGCGGAGGCTTCTCGGTAAACGAGCGGGGTAGCCCGGGGATGGCG 514
QY 253 ProGlyAlaHisIleAlaValTyrLysValCysTyrPheAsnGlyCysTyrSerSerAsp 272
Db CCAGGAGCTCAGTGGCGGTTTCAAGAGTCTCTGCTTCAACGGCTGTTCAGCTCAGAC 454
QY 273 IleLeuAlaAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLysSerLeu 292
Db ATTCTGGCGGATAGACGATAGCGGATACAGCAAGTCCAGCTTCTATCCTCTCCCTA 394
QY 293 GlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMet 312
Db GCGGCTTCCCTATCCTTTGTATGATGACACTATCGCCATTGTGTACCTTCGCGCCACG 334
QY 313 GluArgGlyIleSerValIleCysAlaAlaGlyValAsnAsnGlyProIleGluSerSerVal 332
Db GAACATGGAATCTCCGTTGTCTGCGCGGTGGCAACACGGCCCTCGCTAGCTCGGTG 274
QY 333 AlaAsnThrAlaProIleValSerThrIleGlyAlaGlyThrLeuAspArgArgPhePro 352
Db GCGAACACAGCTCCTTGGCTCTCAACCATCGGAGCAGGAGCTCGATAGAAATTCCTG 214
QY 353 AlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLys 372
Db GGGGTTGTTAGTTAGCCACCGGAAGCTACTCTACGAGAGTCTGTTCACCCAGGAAA 154
QY 373 GlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyLysPlyGly 392
Db GGGATAAAGAGCTCGAGAGAGTGGAGATTGTCTACGTGGCGGAGGAGACAAAGGA 94
QY 393 SerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIle 412
Db AGCGAGTTTGTGTTGAGAGCTCGCTTCGAGAGAGAGTGTCCAAAGGCAAGATGTGTATA 34
QY 413 CysAspArgGlyValAsnGlyArgSerGluLys 423
Db TGGACCGTGGAGTCAACGGGAGATCAGAGAAA 1

RESULT 7

BH442444/c 640 bp DNA linear GSS 12-DEC-2001
LOCUS BOHGT85TR BOHG Brassica oleracea genomic clone BOHGT85, genomic survey sequence.

ACCESSION BH442444
VERSION BH442444.1 GI:17628158
KEYWORDS GSS.

SOURCE Brassica oleracea
ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; euroids II; Brassicales; Brassicaceae; Brassica.

REFERENCE

1 (bases 1 to 640)

AUTHORS

Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other_GSSs: BOHGT85TF
Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES

Location/Qualifiers

1..640
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_lib="BOHGT85"
/clone_lib="BOHG"
/note="Vector: pHS01; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHS01 using BstXI linkers"

BASE COUNT 149 a 211 c 126 g 154 t

ORIGIN

Alignment Scores:

Pred. No.: 5.16e-92 Length: 640
Score: 979.00 Matches: 184
Percent Similarity: 96.1% Conservative: 16
Best Local Similarity: 88.46% Mismatches: 8
Query Match: 24.37% Indels: 0
DB: 28 Gaps: 0

US-09-806-767-2 (1-775) x BH442444 (1-640)

QY 396 CysLeuArgGlySerLeuProArgGluIleArgGlyLysMetValIleCysAspArg 415
Db TGTCTGAGAGCTCGCTCCCGAGAGAGTGTCCAAAGCAAGATGGTGTATGCGACCTG 567
QY 416 GlyValAsnGlyArgSerGluLysGlyAlaValLysGluAlaGlyValAlaMet 435
Db GGAGTCAACGGGAGATCAGAAAGACAGCGGTTAAAGAGCTGGCGCGCTCGCGATG 507
QY 436 IleLeuAlaAsnThrGluIleAsnGlnGluLysAspSerIleAspValHisLeuLeuPro 455
Db ATCTTGCTACCAACCGAGATTACCAAGAGGAGGATTCTGTAGACCTTCTTGTATACCA 447
QY 456 AlaThrIleGlyTyrThrGluSerValLeuLysAlaTyrValAsnAlaThrVal 475
Db GCTACGTTGATTGGTTACGAAGAGTCTGTGTTTGAAGGTTTACGTGAGGACACGGTA 387
QY 476 LysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
Db AGACCGAAGCTAGTTAATTTTCGGCGGAGCGGTGATTTGGGAGGTCGAGAGGCGCTGAG 327
QY 496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515
Db GTGGCGCAGTTCTCAGCGAGACCGAGTTTGGCTTACCCCTACCCCTCGGTGCTTAAACCGGAT 267
QY 516 MetIleAlaProGlyValAsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGly 535
Db TTGATCGCGCAGGTGTCAATATTATAGCTGTCTTGGCTTCAGAACTTCGAGCGCGGG 207
QY 536 LeuProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCys 555
Db CTTCCTTATGATTCGAGGAGAGTTAACTTCACGTGATGTCGGGAGCTTCGATGTTGT 147
QY 556 ProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAla 575
Db CCGCACGTGAGTGGATCACCCTCTCATCGCTCTCGGTATCCGAGCTGTCTCCAGCT 87
QY 576 AlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIle 595

Qy	471	ValAsnAlaThrValIysProIysAlaArgIleIlePheGlyThrValIleGlyArg	490
Db	547	ATCAGTCTCGACGCCAGCGCGTGGCGAGATCGTATTCGGCGGCACGCGATCGGACGA	488
Qy	491	SerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSer	510
Db	487	CGCGCGCGCCCGCGGCGGTGGTCTTCGGCGCGCGCGCGGAGCTTGACCAACCGCTCG	428
Qy	511	IleLeuIysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsn	530
Db	427	GTGCTGAAGCCCGACGCTGTCGCCCCCGGGTGAAATCATCATCGCGCGTGGCCCGGGAAC	366
Qy	531	LeuGlyProThrGlyLeuProTyrAspSerArgValAsnPheThrValMetSerGly	550
Db	367	CTGGCGCGCGTGGCGGCTGGAGAGCAGCGCCCGCGTCCAACTTCACCGTCTCGGGG	308
Qy	551	ThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyPro	570
Db	307	ACATCGATGCGCGCGCTCATGTAGTGGATCGCGCGCTGATCGCGTGGCGCACCCG	248
Qy	571	AsnTrpSerProAlaAlaLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArg	590
Db	247	TCCTGGAGCCAGCGATGTCGGTCCGGATCATGACACCGACTGACATAATCGACCGG	188
Qy	591	GlnGlyIysAlaIleLysAspGlyAsnLysProAlaGly-----ValPheAlaIle	607
Db	187	CAGGGGAAGCGCATCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATG	128
Qy	608	GlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGln	627
Db	127	GCGCGCGCGCACGTGAGCCCGCGCGCGCGTGGACCGCGCTCTGACGATATCCAG	68
Qy	628	ProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIle	645
Db	67	CCGCCGACTACGTGAGCACCTGTGCACGCTCGGGTACAGCCACATGGAGATC	14
RESULT 9			
LOCUS	BH597572	627 bp DNA linear	GSS 15-DEC-2001
DEFINITION	BOHON29TF BOHO Brassica oleracea genomic clone BOHON29, genomic		
ACCESSION	BH597572		
VERSION	BH597572.1		
KEYWORDS	GSS.		
SOURCE	Brassica oleracea		
ORGANISM	Brassica oleracea		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; ;		
AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids		
TITLE	1 (bases 1 to 627)		
JOURNAL	Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.		
COMMENT	Whole genome shotgun sequencing of Brassica oleracea		
	Other GSSs: BOHON29TR		
	Contact: Chris Town		
	TIGR		
	9712 Medical Center Drive, Rockville, MD 20850, USA.		
	Tel: 301-838-3523		
	Fax: 301-838-0208		
	Email: ctown@tigr.org		
	DNA is from a doubled haploid provided by Tom Osborn.		
	Seq primer: TF		
	Class: sheared ends.		
FEATURES	Location/Qualifiers		
source	1..627		
	/organism="Brassica oleracea"		
	/mol_type="genomic DNA"		
	/strain="TO100DH3"		
	/db_xref="taxon:3712"		
	/clone="BOHON29"		
	/clone_lib="BOHO"		
	/note="vector: pHOSt; Site 1: BatXI; 2-3 kb sheared		
	genomic DNA inserted into pHOSt using BstXI linkers"		

Overgo Probes
Unpublished (2002)
2 (bases 1 to 1795)
Coe, E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones, these are publicly available from ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

Location/Qualifiers
1..1795
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:635508"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public configs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 332 a 587 c 603 g 273 t
ORIGIN

Alignment Scores:
Pred. No.: 9,786-82 Length: 1795
Score: 887.50 Matches: 199
Percent Similarity: 58.1% Conservative: 78
Best Local Similarity: 41.8% Mismatches: 174
Query Match: 22.0% Indels: 25
DB: 11 Gaps: 12

US-09-806-767-2 (1-775) x AY108210 (1-1795)

QY 302 AspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAla 321
Db 48 GACCTCATCGCCATCGCCACGCTTCAAGGCCATGAGACGCGGATCTTCGTCAGCGCGCG 107
QY 322 AlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThr 341
Db 108 GCGGGCATAGCGGCGCGCGCGCGGCTCCATCCACACGCGCGCCATGATGATCTCACG 167
QY 342 IleGlyAlaGlyThrLeuAsnArgArgPheProAlaValValArgLeuAlaAsnGlyIys 361
Db 168 GTCCGCGCGCGGACCATGAGACCGCGCATAGCACACCGTCAGGCTCGCGGACGCGGAC 227
QY 362 LeuLeuTyrglyGluSerIleuTyProGlyIysGlyIleIysAsnAlaGlyArgGluVal 381
Db 228 GTGTTGCGAGCGGAGTGTCTGTTCAGCGCGCGGAC---AACACGCGGCGCGCGCGCTG 284
QY 382 GluValIleTyVal---ThrGlyGlyAspIysGlySerGluPheCysLeuArgGlySer 400
Db 285 CGCTCGTCTTCCCGGCGCGACCGCGCGCGGAGCGCGCGGCTGCG---AGCAGC 338
QY 401 LeuProArgGluGluIleArgGlyIysMetValIleCysAsp---ArgGlyValAsnGly 419
Db 339 CTGGTGAGCGCGGAGTGTGAGGCGCAAGGTGTGTGTCTGCGAGAGCGCGCTCCATCAGCG 398
QY 420 ArgSerGluIysGlyGluAlaValIysGluAlaGlyValAlaMetIleLeuAlaAsn 439
Db 399 CACGTGAGAGCGGACGATGTGTCCCGTACGCGCGCGCGCGGATGATCTCATGAAC 458
QY 440 ThrGluIleAsnGluGluAspSerIleAspValHisLeuLeuProAlaThrLeuIle 459
Db 459 AGCCGCGCGGAGGGTTCACACCTTCGCCGAGCGGACGCTCTTGGCGGCTCGCACGTG 518

Alignment Scores:
Pred. No.: 3,776-86 Length: 627
Score: 923.00 Matches: 177
Percent Similarity: 92.34% Conservative: 16
Best Local Similarity: 84.69% Mismatches: 15
Query Match: 22.97% Indels: 1
DB: 28 Gaps: 0

US-09-806-767-2 (1-775) x BH597572 (1-627)

QY 50 LeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSerSer 69
Db 1 CTTTCTTTCTTCAAGAGCTCTTCTAGTATTGAAGAGAAACGAGACCTTCTACA 60
QY 70 ArgLeuLeuTygSerTygSerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSer 89
Db 61 AGGATTCTCTACTTCTACCCCTCTGCCATTTGAAGAGATCTCAGCTCAGTTAACGATCA 120
QY 90 GluAlaGluIleLeuArgTygSerProGluValAlaValAlaValArgProAspHisValLeu 109
Db 121 GAAGCCAAACACTCAGAACTTACTGAGTGTGTGAGTAAGACCTGACCACTGTTCTC 180
QY 110 GlnValGlnThrThrTygSerTygIysPheLeuGlyLeuAspGlyPheGlyAsnSerGly 129
Db 181 CAACTCAAACCACTTACTCTTACAAGTCTTAGGACTCAACGGTCCAGGACCCCTCCAGC 240
QY 130 ValTrpSerIysSerArgPheGlyGlnGlyThrIleGlyValLeuAspThrGlyVal 149
Db 241 GTATGCTCTAAATCCCGGTCGCGCAAGCAGCATCTCGCGCTTCTCGACAGGAGTC 300
QY 150 TrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArgIysTrpIys 169
Db 301 TGCCCCGAAGCCCTAGCTTCACACACAGGAGTGCCTTCAGTCCCAGCAATGGAAA 360
QY 170 GlyIleCysGlnGluGlyGluSerPheSerSerSerCysAsnArgIysLeuIleGly 189
Db 361 GGTGTTTGCAAGAGAGAGAGCTTCACTTCCCTCAACTGTAAACAAGAACTAATCGGC 420
QY 190 AlaArgPhePheIleArgGlyHis-ArgValAlaAsnSerProGluGluSerProAsnMe 209
Db 421 GCTAGATTCTTCATCAGAGCGCCCGGTCAGCAACTCTCTTTGGACTCACCACAAACAT 480
QY 209 tProArgGluTygIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerTh 229
Db 481 GCTCCGGAATCATATCCGACAGAGATCCACGGGGCAGCGGACCCACACCGCTCAAC 540
QY 229 rValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaAr 249
Db 541 TGCTGCGGCTCTCTGTTTCAGCGGAGAGCGTCTCTCGTAACGAGCGCGGTAGCCCG 600
QY 249 gGlyMetAlaProGlyAlaHisIle 257
Db 601 CGGGATGGCGGAGGAGCTCACGTG 625

RESULT 10
AY108210
LOCUS AY108210 1795 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PC0096605 mRNA sequence.
ACCESSION AY108210
VERSION AY108210.1 GI:21211288
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1795)
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitesitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of


```

QY 460 GlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAla 479
Db 519 AGCTACCGCGCGGGTCCAGATCGCGGGGTATACATCAAGTCAGCCAGCGCCAGCGGCG 578
QY 480 ArgIlePheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPhe 499
Db 579 ACATACAGTTCGGGGCACGGTATGGGTCTGCTCCCGCGCGCGTCCGTCGCTCTTC 638
QY 500 SerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaPro 519
Db 639 TCGTCGCGCGCGCGAACAGGCGAGCGCGGCATCTCTGAAGCCGACATCACCGGGGCC 698
QY 520 GlyValAsnIleLeuAlaLafp---ProGlnAsnLeuGlyProThrGlyLeuProTyr 538
Db 699 GGAATGAACATCTCGCGCGGTGGCGCCGAGCGAGATGACCCGAGTTCGCG----- 752
QY 539 AspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisVal 558
Db 753 GACGAGTGTAGCTGCCCTTCTCATGGAGTCCGGGACGTCATGTCGAGCGCGCACCTG 812
QY 559 SerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTyrSerProAlaAlaIleLys 578
Db 813 AGCGGATCGCGGCATCATCAGAGCCTGCACCGAGTGTGTCCTCCCGCGCGATCAAG 872
QY 579 SerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLeuAlaIleLysAspGly 598
Db 873 TCGGCGATCATGACGTCGTCGCGCACGCGCGGACCGCGGCGTGCAGATCAAGACGAG 932
QY 599 Asn---LysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAla 617
Db 933 CAGTACAGCGCGCGAGCTTACAGCATGGCGCGCGCTAGCTGAACCGCTCCGCGCGC 992
QY 618 IleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThr 637
Db 993 GTTCACCGCGCGCTGTGTACACCTCGCGCGCGGAGTACATCGCTACCTCTCGCGGG 1052
QY 638 LeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGly 657
Db 1053 CTGGGATCGGGACGCGCGTGAAGGAGATACTGGCGCGCGCTGCGCTGC---GCC 1109
QY 658 IleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLys 677
Db 1110 AAGCTGAAGGCCATCACCGAGCGGAGCTCAACTACCGCTCGCTGTGTGAAGTGCTG 1169
QY 678 ArgGlyValThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSer 697
Db 1170 TCCATPCGATCACC-----GTCCCGCGCACCGTGAACCAAGTGGGAAGGCCAATCG 1223
QY 698 IleTyrSerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArg 717
Db 1224 GTGTACAGCGCGTGTGTGACATGCCAGCGCGGTGTCGCTGCGTGGTGGCGCCCGGTG 1283
QY 718 LeuValPheLysHisValAspGlnThrLeuSerTyrArgVal-----TrpPheVal 734
Db 1284 CTGCGTTCGCCAGCGCGAACGAGAGAGAGCTTACCGTACCGTGGCGGTG----- 1337
QY 735 LeuLysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGlyClnLeuThrTrp 754
Db 1338 -----AACGGCGCGCGCGCGTGGCGCGCGCGCGCGAGGAGAACCTGAAGTGG 1382
QY 755 ValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal 770
Db 1383 GTGTGAGGAGCACGTC-----GTCCGAGCGCGCGATCGTTATA 1421

RESULT 11
BH601308/c 803 bp DNA linear GSS 15-DEC-2001
LOCUS BOHSA897F BOHS Brassica oleracea genomic clone BOHSA89, genomic
DEFINITION survey sequence.
ACCESSION BH601308
VERSION BH601308.1 GI:17853754
KEYWORDS GSS.
SOURCE Brassica oleracea

```

```

ORGANISM Brassica oleracea
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
TITLE euroids II; Brassicales; Brassicaceae; Brassica.
JOURNAL 1 (bases 1 to 803)
COMMENT Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BOHSA89TR
Contact: Chris Town
TIGR Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
source
1..803
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHSA89"
/clone_lib="BOHS"
/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 245 a 223 c 121 g 214 t
ORIGIN
Alignment Scores:
Pred. No.: 1,09e-79 Length: 803
Score: 863.00 Matches: 167
Percent Similarity: 91.13% Conservative: 18
Best Local Similarity: 82.27% Mismatches: 16
Query Match: 21.48% Indels: 2
DB: 2 Gaps: 2
US-09-806-767-2 (1-775) x BH601308 (1-803)
QY 575 AlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAla 594
Db 803 GCTGCTATCAGATCAGCGATGATGACGCGCTGATTGTGATGATCGGAGGGAAGAG 744
QY 595 IleLysAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro 614
Db 743 ATTAGGACGCGGACAAACCGCGGAGTGTGCTATTGGACGCGCATGTGAATCCG 684
QY 615 GlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyr 634
Db 683 GTTAAGCGGATTAACCGGGGTGGTTTACACATCAACCGGTTGATTGCTTAC 624
QY 635 LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSer 654
Db 623 TTATGCACTATTGGTTTACTAGATCGATATTATTAGCATCTACTCAGGAACGTTAGC 564
QY 655 CysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaVal 674
Db 563 TCGGTCCGCTACTAGGAGAGCCCGCGTTTGTAGTCTTAATTCGCTATTTCGTT 504
QY 675 IlePheLysArgGlyLysThrGluMetIleThrArgArgValThrAsnValGlySer 694
Db 503 ATTTTAGCGTGGGAGGACTAAGGAGATGATCAGGAGCGGTGTGACTAACGTTGGAGT 444
QY 695 ProAsnSerIleTyrSerValAsnValLysAlaProGluGlyLysValIleValAsn 714
Db 443 CCTAAGTCCGTGTACACTGTGAATGTAAGCTCCCTATGGGGATTATGTGATTGTAAG 384
QY 715 ProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTyrPheVal 734
Db 383 CCTAAGAGGCTTGTGTTAGTACGTTGATCAACAGCTTGAGCTATAGAGTTGCTTGTG 324

```


REFERENCE AUTHORS	TITLE	JOURNAL
1 (bases 1 to 767) Tomkins, J. P., Fregene, M., Main, D., Goicoechea, J. L., Blackmon, B., Atkins, M., Tohme, J. and Wing, R. A.	New Genomic Resources for Cassava (Manihot esculenta): Development of a Deep-Coverage BAC Library and Preliminary STC Analysis	Unpublished

CONTACT: Tomkins J
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 6419
Fax: 864 656 4293
Email: jtmkns@clemson.edu
Total High Quality bases = 313
Seq primer: TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence start: 77
High quality sequence stop: 546

```

FEATURES
source
    Location/Qualifiers
        1..767
            /organism="Manihot esculenta"
            /mol_type="genomic DNA"
            /strain="MECW72"
            /db_xref="taxon:3983"
            /clone="ME_MBa0001K2zf"
            /tissue_type="Leaf"
            /lab_host="E. coli"
            /clone_lib="Manihot esculenta"
            /note="Vector: pCUGIBAC-1; Site 1: HindIII; Site 2: NotI;
            For more details on library preparation and sequence
            analysis see
            http://www.genome.clemson.edu/projects/stc/cassava/ME_MBa
            to order clones from this library see
            http://www.genome.clemson.edu/orders "
233 a 183 c 153 g 186 t 12 others

```

Alignment Scores:		
Pred. No.:	1.74e-70	Length:
Score:	775.00	Matches:
Percent Similarity:	78.93%	Conservative:
Best Local Similarity:	65.70%	Mismatches:
Query Match:	19.23%	Indels:
DB:	28	Gaps:
		0
US-09-806-767-2 (1-775)	x	BH793726 (1-767)

Db	405	TCAGCAATAATGACTACTGCAGATATAACTGACCACTTCGGAATCAATATTGGATGGT	464
Qy	599	ASnLysProAlaGlyValPheAlaIleGlyVala-GlyHisValAsnProGlnLysAlaIle	618
Db	465	GACAACACGACGACATCTTGGCAATTGGAGCTGCCCATATACACCCAGAGAGGCAT	524
Qy	618	eAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLe	638
Db	525	CAATCCGAGATTGATATATGATATCAGGCCAGAGAGTATGTCACCCACCTATGNCACCC	584
Qy	638	uGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIle	658
Db	585	TAGATACACAGATCAGAAATTTTCAATATCACACAGAGATGTGAGCTGNCATGAAC	644
Qy	658	eLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysAr	678
Db	645	TATGCGAGATGAATAGGTTNCAGCTTCAATATTCCTTNCATTCCTGTTATNTAAAGCG	704
Qy	678	gGlyLysThrThrGluMetIleThrArgValThrAsnValGlySerProAsnSerIle	698
Db	705	GGGATG-ACNNAGTAGATGATCCGAAGACGACTTACCAATGTGGTAT-CCTAAATCAT	762
Qy	698	eTyr 599	
Db	763	CTAC 766	
RESULT 14			
BP052773			
LOCUS	BP052773	568 bp	linear EST 07-MAR-2003
DEFINITION	EST438003 potato leaves and petioles Solanum tuberosum cDNA clone		
ACCESSION	CSB32P4	5' sequence, mRNA sequence.	
VERSION	BP052773		
KEYWORDS	BP052773.1	GI:10806669	
SOURCE	EST.		
ORGANISM	Solanum tuberosum (potato)		
REFERENCE	Solanum tuberosum		
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
COMMENT	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	asterids; lamids; Solanales; Solanaceae; Solanum.		
	1 (bases 1 to 568)		
	van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,		
	Utterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Renning,		
	C.M., Fry,W.E., Tanksley,S.D. and Baker,B.		
	Generation of ESTs from potato leaves and petioles		
	Unpublished		
	Contact: Robin Buell		
	The Institute for Genomic Research		
	9712 Medical Center Dr, Rockville, MD 20850, USA		
	Email: potato-array@tigr.org		
	This clone can be obtained from the University of Arizona Genomics		
	Institute. Orders can be made through URL:		
	http://genome.arizona.edu/orders/.		
FEATURES	Location/Qualifiers		
	1..568		
	source		

```

/organism="Solanum tuberosum"
/mol_type="rRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="CST332P4"
/tissue_type="leaflets and petioles"
/dev_stage="8 weeks old plants"
/lab_host="SOLR"
/clone_lib="potato leaves and petioles"
/notes="vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Tissue was supplied by Dr. Fry (Cornell University).
Leaflets and petioles were isolated from 8 week old
greenhouse grown plants. The plants were watered and
fertilized freely. The tissue was immediately frozen in
liquid nitrogen."
176 a 123 c 133 g 136 t
BASE COUNT
ORIGIN
Alignment Scores:

```

Alignment Scores:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 09:06:22 ; Search time 46 Seconds
(without alignments)

2674.198 Million cell updates/sec

US-09-806-767-2

4018

Perfect score:

Sequence: 1 MEKPFPLCIIFLLPCSSSS.....NSHNLQVRSPISVTLKTN 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum first 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4018	100.0	775	21	AAV83300 SDDI subtilisin li
2	4018	100.0	775	23	ABV90836 Herbicidally activ
3	2786.5	69.4	766	21	AAV83304 Subtilase (SC_P2sc
4	2727.5	67.9	829	21	AAV83302 Subtilase (Pigesp)
5	2727.5	67.9	829	21	AAV83303 Subtilase (Pigesp)
6	2540	63.2	491	21	AAV83301 Prematurely termin
7	1742	43.4	754	21	AAG32219 Arabidopsis thalia
8	1742	43.4	754	23	ABV91728 Herbicidally activ
9	1700	42.3	775	21	AAG39671 Arabidopsis thalia

10	1684	41.9	775	21	AAG28659 Arabidopsis thalia
11	1666.5	41.5	757	23	ABV93985 Herbicidally activ
12	1642.5	40.9	780	21	AAG43556 Arabidopsis thalia
13	1642.5	40.9	780	23	ABV93785 Herbicidally activ
14	1612	40.1	735	21	AAG43558 Arabidopsis thalia
15	1612	40.1	740	21	AAG43557 Arabidopsis thalia
16	1589	39.5	764	21	AAG29250 Arabidopsis thalia
17	1575.5	39.2	728	21	AAG29251 Arabidopsis thalia
18	1562	38.9	757	23	ABV90803 Herbicidally activ
19	1526.5	38.0	606	21	AAG32220 Arabidopsis thalia
20	1409.5	35.1	568	21	AAG32221 Arabidopsis thalia
21	1369	34.1	763	23	ABV93893 Herbicidally activ
22	1348.5	33.6	581	21	AAG39672 Arabidopsis thalia
23	1337.5	33.3	581	21	AAG28660 Arabidopsis thalia
24	1309	32.6	791	23	ABV93713 Herbicidally activ
25	1271.5	31.6	756	23	ABV92882 Arabidopsis thalia
26	1248.5	31.1	537	21	AAG39673 Arabidopsis thalia
27	1241	30.9	736	23	ABV93977 Herbicidally activ
28	1237.5	30.8	537	21	AAG28661 Arabidopsis thalia
29	1219.5	30.4	780	23	ABV91074 Herbicidally activ
30	1215.5	30.3	775	23	ABV92884 Arabidopsis thalia
31	1214	30.2	769	21	AAG28896 Arabidopsis thalia
32	1201.5	29.9	748	21	AAG28897 Arabidopsis thalia
33	1201.5	29.9	748	23	ABV91075 Herbicidally activ
34	1200.5	29.9	730	21	AAG28898 Arabidopsis thalia
35	1196	29.8	803	23	ABV92885 Herbicidally activ
36	1190.5	29.6	747	23	ABV92883 Arabidopsis thalia
37	1187.5	29.6	542	21	AAG29252 Arabidopsis thalia
38	1187.5	29.6	706	23	ABV92784 Herbicidally activ
39	1180.5	29.4	736	23	ABV92568 Herbicidally activ
40	1176	29.3	739	23	ABV92567 Herbicidally activ
41	1173	29.2	765	23	ABV92881 Herbicidally activ
42	1172.5	29.2	731	15	AAG60822 Melon protease fro
43	1145.5	28.5	746	23	ABV93111 Herbicidally activ
44	1117	27.8	772	23	ABV93038 Herbicidally activ
45	1101.5	27.4	755	23	ABV93440 Herbicidally activ

ALIGNMENTS

RESULT 1

AAV83300

ID AAV83300 standard; Protein; 775 AA.

XX

AC AAV83300;

XX

DT 16-AUG-2000 (first entry)

XX

DE SDDI subtilisin like serine protease.

XX

KW SDDI; serine protease; subtilisin; transgenic plants; dry weight;

KW stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;

KW crop protection; feed; foodstuffs.

XX

OS Arabidopsis thaliana.

XX

PN WO200022144-A2.

XX

PD 20-APR-2000.

XX

PF 12-OCT-1999; 99WO-EP07633.

XX

PR 12-OCT-1998; 98EP-0119244.

XX

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.

XX

PI Berger D, Altmann T;

XX

DR WPI: 2000-317995/27.

DR

DR N-PSDB; AAV93808.

XX

PT	Novel recombinant DNA molecules encoding subtilisin-like serine protease, useful for producing transgenic plants with altered stomata, lower water consumption and enhanced diseased resistance
PS	Claim 1; Page 63-65; 101pp; English.
PX	
XX	
CC	Sequences encoding SDD1, a subtilisin-like serine protease, can be used to produce transgenic plants with altered stomata characteristics. These plants exhibit improved freshness, increased dry weight, reduced leaf temperatures, reduced water loss and lower water consumption and for enhancing the sugar and/or protein content of plant leaves, modulating CO ₂ uptake into and H ₂ O release from leaves, for sustained photosynthesis under high intensity conditions or for the improvement of disease resistance of plants. The transgenic plants and cells of such plants are useful in the preparation of feed, food or additives.
CC	
XX	
SQ	Sequence 775 AA;
	Query Match 100.0%; Score 4018; DB 21; Length 775;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MEPKPFFLCIIILLFCSSSSILQKQTVIVQLHNPSETAKTFASKDWHLSFLOEAVLGV 60
DB	1 MEPKPFFLCIIILLFCSSSSILQKQTVIVQLHNPSETAKTFASKDWHLSFLOEAVLGV 60
QY	61 EEEEEPSSRLLYSGSAIEGFAAQLTESEAEILURYSPEVVAVRPHVLQVQTTSYKFL 120
DB	61 EEEEEPSSRLLYSGSAIEGFAAQLTESEAEILURYSPEVVAVRPHVLQVQTTSYKFL 120
QY	121 GLDGFGNSGWMSKSIFGGQTTIIGVLDTGWPSPSPDTCMPISIRKWKGICQEGESFSS 180
DB	121 GLDGFGNSGWMSKSIFGGQTTIIGVLDTGWPSPSPDTCMPISIRKWKGICQEGESFSS 180
QY	181 SSCNRKLIGARFFIRGHVRANSPESPNNPREYISARDSTGCHGTHTASTVGSSVSVMNV 240
DB	181 SSCNRKLIGARFFIRGHVRANSPESPNNPREYISARDSTGCHGTHTASTVGSSVSVMNV 240
QY	241 LNGAGVARGMAPGAHAIVYKVCFNCGCYSDIIAAIDVAIQKDVLISLGGRPIPLY 300
DB	241 LNGAGVARGMAPGAHAIVYKVCFNCGCYSDIIAAIDVAIQKDVLISLGGRPIPLY 300
QY	301 DDTIATGTFRAMERGISVICAAGNGPIESSVANTAPWVTICAGTLDRRFPAVVLANG 360
DB	301 DDTIATGTFRAMERGISVICAAGNGPIESSVANTAPWVTICAGTLDRRFPAVVLANG 360
QY	361 KLLYGSLYPGKGIKNAGREVEVIYVTGGDKGSEFCULRGSLPREEIRGMVICDRGNWR 420
DB	361 KLLYGSLYPGKGIKNAGREVEVIYVTGGDKGSEFCULRGSLPREEIRGMVICDRGNWR 420
QY	421 SEKGEAVKAGGVAMILANTEINQBEDSIDVHLIPATLIGYTSSVLLKAYVNAVTKPKAR 480
DB	421 SEKGEAVKAGGVAMILANTEINQBEDSIDVHLIPATLIGYTSSVLLKAYVNAVTKPKAR 480
QY	481 IIFGGTVCGRRAPEVAOFSARGPSLANPSILKDMTAPGVNIITAAPQNGLPTGLPYDS 540
DB	481 IIFGGTVCGRRAPEVAOFSARGPSLANPSILKDXMTPAGVNIITAAPQNGLPTGLPYDS 540
QY	541 RRVNFTWMSGTSMSCPHYSGITALIRSAYPNWSPPAAIKSALMTTADLYDRQGKAIDGKNK 600
DB	541 RRVNFTWMSGTSMSCPHYSGITALIRSAYPNWSPPAAIKSALTADLYDRQGKAIDGKNK 600
QY	601 PACVFPAIGHAVNPQKAINPGLVYNIQPDVITYLCTILGTRSRDILAITHKNVSCNGILR 660
DB	601 PACVFPAIGHAVNPQKAINPGLVYNIQPDVITYLCTILGTRSRDILAITHKNVSCNGILR 660
QY	661 KNPGPSLNVPSTIAVIFKRGKTTEMITRRVTNVGSPNSTIYSVNVKAPSGIKVINPKRLVY 720
DB	661 KNPGPSLNVPSTIAVIFKRGKTTEMITRRVTNVGSPNSTIYSVNVKAPSGIKVINPKRLVY 720
QY	721 KHVDQDTLSYRVNFVLKKRNRGKVASFAQQGLTWVNSHNLMQRVRSPISVTLKTN 775

```

Db      721 KHYDQTILSRVWFVLKKKRRGGKVASFQAQQLTWNASHNLQWRVRSPISVTLKTN 775

RESULT 2
ABB90836
ID      ABB90836 standard; Protein; 775 AA.
XX
XX      ABB90836;
XX
XX      31-MAY-2002 (first entry)
XX
XX      Herbicidally active polypeptide SEQ ID NO 47.
XX
XX      Herbicidal; plant; agriculture; herbicide.
XX
XX      Arabidopsis thaliana.
XX
XX      WO200210210-A2.
XX
XX      07-FEB-2002.
XX
XX      28-AUG-2001; 2001WO-EP09892.
XX
XX      28-AUG-2001; 2001WO-EP09892.
XX
XX      (FARB ) BAYER AG.
XX
XX      Tietjen K, Weidler M;
XX
XX      WPI; 2002-269010/31.
XX
XX      Identifying plant target proteins for herbicidally active compounds,
PT      comprising aligning and comparing nucleic acid or amino acid sequences
PT      from plant with nucleic acid or amino acid sequences from non-plant
PT      organisms -
XX
XX      Claim 5; SEQ ID NO 47; 261bp + Sequence Listing; English.
XX
XX      The invention relates to identifying target proteins
XX      (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX      aligning and comparing nucleic acid or amino acid sequences from plant
XX      with nucleic acid or amino acid sequences from non-plant organisms using
XX      suitable search parameters, where plant sequences having an E-value
XX      greater by a factor of 3 than the E-value of most similar non-plant
XX      sequences are selected. The polypeptides or nucleic acids encoding them
XX      are useful for identifying modulators. The identified modulators are
XX      useful as herbicides.
XX
XX      Query Match          100.0%; Score 4018; DB 23; Length 775;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 MEKPPFLCIIIFLLFCSSESSEILKQYIYVLHPNSETAKTFASKFDWHL-SFLQEAVLGV 60
Db      1 MEKPPFLCIIIFLLFCSSESSEILKQYIYVLHPNSETAKTFASKFDWHL-SFLQEAVLGV 60
Qy      61 EEEEEPSRLLYSGAIGFAAQLTESAEILRYSPEVVAVRDPDVLQVQTYSYKFL 120
Db      61 EEEEEPSRLLYSGAIGFAAQLTESAEILRYSPEVVAVRDPDVLQVQTYSYKFL 120
Qy      121 GLDGFNGSVWKSFRFCGGTII GVLDTGVWPESPFDTCGWPSIPRKWKGICQEGESFSS 180
Db      121 GLDGFNGSVWKSFRFCGGTII GVLDTGVWPESPFDTCGWPSIPRKWKGICQEGESFSS 180
Qy      181 SSCNRKLIIGARFFIRGHRVANSPESENMPREYISARDSTGCHGHTASTVGGSSVSMANV 240
Db      181 SSCNRKLIIGARFFIRGHRVANSPESENMPREYISARDSTGCHGHTASTVGGSSVSMANV 240
Qy      241 LGNGAGVARGVAPGAHIVYKVCWFNCGYSSDILAAIDVAIQDVLSLSLGGFPPILY 300
Db      241 LGNGAGVARGVAPGAHIVYKVCWFNCGYSSDILAAIDVAIQDVLSLSLGGFPPILY 300

```



```

RESULT 4
AAY83302
ID AAY83302 standard; Protein; 829 AA.
XX
AC AAY83302;
XX
DT 16-AUG-2000 (first entry)
XX
DE Subtilase (Pigesp) of Solanum tuberosum.
XX
KW SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomach; sugar; water; protein; CO2; H2O; CO2; H2O;
KW crop protection; feed; foodstuffs; ss.
XX
OS Solanum tuberosum.
XX
Key Location/Qualifiers
FT Misc-difference 225
FT /note= "Unidentified amino acid"
XX
WO200022144-A2.
XX
20-APR-2000.
XX
12-OCT-1999; 99WO-EP07633.
XX
12-OCT-1998; 98EP-0119244.
XX
(PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
Berger D, Altmann T;
XX
WPI; 2000-317995/27.
DR N-PSDB; AA293812.
XX
Novel recombinant DNA molecules encoding subtilisin-like serine
PT protease, useful for producing transgenic plants with altered stomacha,
PT lower water consumption and enhanced diseased resistance
XX
Claim 1; Page 76-78; 101pp; English.
XX
Sequences encoding SDD1, a subtilisin-like serine protease, can be
CC used to produce transgenic plants with altered stomacha
CC characteristics. These plants exhibit improved freshness,
CC increased dry weight, reduced leaf temperatures, reduced water loss
CC and lower water consumption and for enhancing the sugar and/or
CC protein content of plant leaves, modulating CO2 uptake into and H2O
CC release from leaves, for sustained photosynthesis under high
CC intensity conditions or for the improvement of disease resistance
CC of plants. The transgenic plants and cells of such plants are useful
CC in the preparation of feed, food or additives. This subtilase of
CC Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
CC (See AAY83300).
XX
SQ Sequence 829 AA;
Query Match 67.9%; Score 2727.5; DB 21; Length 829;
Best Local Similarity 65.8%; Pred. No. 4e-222;
Matches 521; Conservative 108; Mismatches 132; Indels 33; Gaps 5;
XX
6 FFLCIIIFLPCSSSSBILOKQYIVQLHPNSETAKT-FASKFDWHLSTLQEQ----- 55
34 YFLC--FLLCIFLLQANLQYIVQLHPQHASTRIPFSKQWHLSTLENITNIPNF 91
56 -----AVLGEVEEPEPSRRLLYSGSAIEGFAAQLTSEAEILRY 96
92 KYIOWNSIPILFCFYSVSPATSISSGENSSRLLYSYHSAFEGFAALLSENELKALKK 151
97 SPVAVRPHVQLVQVTTYSYKFLGLDGFNGSNVWSKSRFGQGTIGVLDGTWBPSPSF 156

```

```

Db 152 SNNVLSIYPERKLEVVQTTYSYKFLGLSP-TKEGTWLKSGFGRGAILGVLDGTWBPSPSF 210
Qy 157 DDTGMPISIPKWKIGIQEGESFSSSSCNKRLICARFFIRGHRVANSPEESPNNPREYISA 216
Db 211 VDHGMSPIPKWKIGXQEGKNFSSCNKRLICARFFQIGHMMASKTSKSIDFMEDYVSP 270
Qy 217 RDSGTGHTTASTVGGSSVSMANVLNGAGVARGMAPGAHIAVYKVCWNGCYSSDILAA 276
Db 271 RDSQGHGHTTASTAGGAPVPMASVLNGAGARGMAPGAHIAIYKVCWSSGCYSSDILAA 330
Qy 277 IDVAIQKVDVLSLSLGGPPIPLYDITIAIGTFRAMEGIVSICAAGNNGPIESSVANITA 336
Db 331 MDVAIRDGVLDLSLSIGGFPVLYEDTIIAGSFRAMERGLSVICAAGNNGPILUSSVANEZ 390
Qy 337 PWSTTIGAGTLDRRFFAVVRLANGKLLYGESLYPGKIKNAGREVIVYVTGDKGSEFC 396
Db 391 PWIATIGASTLDRKFPALIQLGNGKYVGESLYPGKOVENSOKVLEIVLNDGNGSEFC 450
Qy 397 LRGLSPREEIRGKWIICDRGVNGRSEKGEAVKEAGGVAMILANTEINQEDSIDVHLLPA 456
Db 451 LRGLSPRAKVHGKIIVCDRGVNGRAEKGVVKESSGGVAMILANTAVNMEEDSVDVHLEPA 510
Qy 457 TLIGYTESVLLKAYVNATVVKPARIIFGGTVIGRSPAPEVAQESARGPSLANPSILKPD 516
Db 511 TLIGFDESIQLOQSYMNSTRKPTARIIFGGTVIGKSSAPAVAQFSSRGPSFTDPSILKPDV 570
Qy 517 IAPGVNIIAAMPONLPGTGLPYDSRVNFTVMSGTSMSCPHVSGITALIRSAVPNWSPPAA 576
Db 571 IAPGVNIIAAMPONLPGSLAEDSRVNFVLSGTSMACPHVSGIAALHSHHPKWSPPAA 630
Qy 577 IKSALMTTADLYDRQKKAIDGNKPKAGVFAIGAAGHVNPQKAINPGLVYNIQVDTYITLC 636
Db 631 IKSALMTTADTTHQKPFIMDGTAGLFAIGAAGHVNPGRSDDPGLIYDINANDYITHLC 690
Qy 637 TLGFTSRDILAIATHKNVSCNGILRKNPGLPSLNPSTAVIFKRGKTEMTTRRTVNVGSPN 696
Db 691 TIGYKXSEILSIHKVNSCHDVLOKVRGSLNPSISVIFKAGKTRKMITRRTVNVGSPN 750
Qy 697 SIYSVNVKAPEGKVIIVNPKRLYKHVDQTLVSRVNFVLKKNRGKGVASFAOGLTWN 756
Db 751 SIYSVEIVAPEGVKVRKPRRLVFKHVNOSLSYRVWF1-SRKIGIGTQRSSFAEGQLMWIN 809
Qy 757 SHNLMQVRSPISV 770
Db 810 SRDKYQKVRSPISV 823
RESULT 5
AAY83303
ID AAY83303 standard; Protein; 829 AA.
XX
AC AAY83303;
XX
DT 16-AUG-2000 (first entry)
XX
DE Subtilase (Pigesp) of Solanum tuberosum.
XX
KW SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomach; sugar; water; protein; CO2; H2O; CO2; H2O;
KW crop protection; feed; foodstuffs; ss.
XX
OS Solanum tuberosum.
XX
Key Location/Qualifiers
FT Misc-difference 225
FT /note= "Unidentified amino acid"
XX
WO200022144-A2.
XX
20-APR-2000.
XX
12-OCT-1999; 99WO-EP07633.
XX

```


PR 12-OCT-1998; 98EP-0119244.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
PI Berger D, Altmann T;
XX
DR WPI; 2000-317995/27.
DR N-PSDB; AAZ93813.
XX
PT Novel recombinant DNA molecules encoding subtilisin-like serine
PT protease, useful for producing transgenic plants with altered stomata,
PT lower water consumption and enhanced diseased resistance
XX
PS Claim 1; Page 82-84; 101pp; English.
XX
XX Sequences encoding SDD1, a subtilisin-like serine protease, can be
CC used to produce transgenic plants with altered stomata
CC characteristics. These plants exhibit improved freshness,
CC increased dry weight, reduced leaf temperatures, reduced water loss
CC and lower water consumption and for enhancing the sugar and/or
CC protein content of plant leaves, modulating CO₂ uptake into and H₂O
CC release from leaves, for sustained photosynthesis under high
CC intensity conditions or for the improvement of disease resistance
CC of plants. The transgenic plants and cells of such plants are useful
CC in the preparation of feed, food or additives. This subtitase of
CC Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
CC (See AA83300).
XX
XX Sequence 829 AA;
SQ
Query Match 67.9%; Score 2727.5; DB 21; Length 829;
Best Local Similarity 65.6%; Pred. No. 4e-222;
Matches 521; Conservative 108; Mismatches 132; Indels 33; Gaps 5;
6 FFLCIIFLCSSSSBILOKQYIVQLHPNSHTAKT-FASKEDWHLFLQE----- 55
34 YFLC--FLLCFTPLLAQNLQYIVQLHPHASTRTFPSSKQWHLSPLENTNPLNF 91
56 -----AVLGEDEEPEPSRLLYSGSAIEGFAAQLTESEAIIRY 96
92 KYIQWNSIFILFCFYSVPATSISSGENSSRLLYSYHSAFEGFAALLSENELKALK 151
97 SPBVAVRPHVLQVCTTYSYKFLGDDGNSGVWSKRFGGTIIIGVLDGWPEPSPF 156
152 SNNVLSIYPERKLEVTYISYKFLGSP-TKSGTWXSGFGGAIIGVLDTIWPEPSPF 210
157 DDTGMPISIPKWKIGICQEGESFSSSCNKLICARFFIRGHRVANSPEPSNMPREYISA 216
211 VDHGMSPIPKWKXGQCGENFNSSCNKLICARFFQIGHNMASKTSIDFMDYVSP 270
217 RDSGTGHTASTVGGSSVSMANVLGNAGVARGMAPGAHIAVYKCVWFGYSSDILAA 276
271 RDSQGHGTHTAGGAPVPMASVLGNAGARGMAPGAHIAIYKCVSSGCGYSSDILAA 330
277 IDVAIQKVDVLSLGGPPIPLDYDTIAGTFRAMERGISVICAGNNGPIESSVANTA 336
331 MDVAIRDGVLDLSLGGPFPVLYEDTIAIGSFAMERGISVICAGNNGPITLSSVANE 390
337 PWYSTTICAGTLDRRFPVAVRLANGKLLYGESIYVPGKIKNAGREVEIVVTGDDGSEFC 396
391 PWATIGASTLDRKFFAIQLGNGKIVYGESIYVPGQVHNSQKVIIEVLNDGNGSEFC 450
397 LRGLSPREIRGHWICDRGVNRSEKGAIVKAVAGGVAMILANTINOEDSIDVHLLPA 456
451 LRGLSPRAKHGKIVVCDRGVNGRAEKGVVXESGGVAMILANTAVNNEEDSDVHLLPA 510
457 TLIGYTESVLLKAYNATVKPARIIFGQTVIGRRAPEVAQFARGSPSLANPSILKPD 516
511 TLIGFDESILQLYNMSTRKPTARIIFGQTVIGKSAVAQFSGRGSFTDPSILKPDV 570
517 IAPGVNIIAAMPONLGTGLPYDSRVNFTVMSGTSMSCPHVSGITALIRSAFVNSPAA 576

Db 571 IAPGVNIIAAMPONLGPGLAEDSRVRNFTVLSGTSMACDHVSGIAALLHSIHPKWPAA 630
Qy 577 IKSALMTTADLYORQKAIKIDGNKPAIGFAIGHVNPQKAINPGLVYNIQPYDYITLC 636
Db 631 IKSALMTTADTNHOGKPIINDGTRAGLFAIGHVNPQSGDDPGLIYDINANDYITHLC 690
Qy 637 TLGFTSRDILAIATHKNVSCNGIIRKQDPGFSNLTPSIAVIPKRGKTTMTTRVTNVGSPN 696
Db 691 TIGYKNSIILSIHKNVSCNDVLQKRGFSNLTPSISVIPKAGTRKMTTRVTNVGSPN 750
Qy 697 SIYSVNVKAPBGLKIVNPKLVFKVDOTLSYRVNFKKNGGKVASFAOGLTWN 756
Db 751 SIYSVEIVABEGVKVRKPRRLVFKVNQSLSYRVNFI-SRKRIGTQRSSFAEQMLWIN 809
Qy 757 SHNLMQVRSPISV 770
Db 810 SRDKYQKVRSPISV 823
RESULT 6
AA83301
ID AA83301 standard; Protein; 491 AA.
XX AA83301;
AC AA83301;
XX
DT 16-AUG-2000 (first entry)
XX Prematurely terminated SDD1 subtilisin like serine protease.
DE
XX SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;
KW crop protection; feed; foodstuffs.
XX Arabidopsis thaliana.
XX WO200022144-A2.
XX 20-APR-2000.
PF 12-OCT-1999; 99WO-EF07633.
XX
PR 12-OCT-1998; 98EP-0119244.
XX
(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
(PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
Berger D, Altmann T;
WPI; 2000-317995/27.
N-PSDB; AAZ93809.
Novel recombinant DNA molecules encoding subtilisin-like serine
protease, useful for producing transgenic plants with altered stomata,
lower water consumption and enhanced diseased resistance
Claim 1; Page 68-70; 101pp; English.
Sequences encoding SDD1, a subtilisin-like serine protease, can be
used to produce transgenic plants with altered stomata
characteristics. These plants exhibit improved freshness,
increased dry weight, reduced leaf temperatures, reduced water loss
and lower water consumption and for enhancing the sugar and/or
protein content of plant leaves, modulating CO₂ uptake into and H₂O
release from leaves, for sustained photosynthesis under high
intensity conditions or for the improvement of disease resistance
of plants. The transgenic plants and cells of such plants are useful
in the preparation of feed, food or additives. This sequence is
identical to the one given in GENESEQ record AA83300 with the
exception that a single point mutation in the coding sequence
introduces a TGA codon and results in premature termination
of the SDD1 polypeptide.
XX Sequence 491 AA;
SQ

```
Query Match          63.2%; Score 2540; DB 21; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.4e-206;
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEPKPFLCIIIFLLCCSSSEILQQTIVIVQLHPNSETAKTFASKFDWHLFLQEAVLGV 60
Db 1 MEPKPFLCIIIFLLCCSSSEILQQTIVIVQLHPNSETAKTFASKFDWHLFLQEAVLGV 60
QY 61 EEEEEPSSRLIYSGSAIEGFAAOLTESEABILYSEVAVRDPDHVLQVTTYSYKFL 120
Db 61 EEEEEPSSRLIYSGSAIEGFAAOLTESEABILYSEVAVRDPDHVLQVTTYSYKFL 120
QY 121 GLDGFNSGVNSKSFQGTIIIGVLDTGWPSPSFDGTGMPSPRKKWGIQQGESEFSS 180
Db 121 GLDGFNSGVNSKSFQGTIIIGVLDTGWPSPSFDGTGMPSPRKKWGIQQGESEFSS 180
QY 181 SSCNRKLIARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHASTVGGSSVSMANV 240
Db 181 SSCNRKLIARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHASTVGGSSVSMANV 240
QY 241 LONGAGVAGMAPAGAHIAVYKVCWNGCYSSDILAAIDVAIQDKVDVLSLSLGGFPPIPLY 300
Db 241 LONGAGVAGMAPAGAHIAVYKVCWNGCYSSDILAAIDVAIQDKVDVLSLSLGGFPPIPLY 300
QY 301 DDTIAGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGACTLDRRPPAVVRLANG 360
Db 301 DDTIAGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGACTLDRRPPAVVRLANG 360
QY 361 KLLYGESLYPGKIGKNAGREVEVIYVTGDKGSEFCLRGSLPREIRKMWICDRGVNGR 420
Db 361 KLLYGESLYPGKIGKNAGREVEVIYVTGDKGSEFCLRGSLPREIRKMWICDRGVNGR 420
QY 421 SEKGEAVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGTYESVLLKAYVNATVKKPKAR 480
Db 421 SEKGEAVKEAGGVAMILANTEINQBEDSIDVHLLPATLIGTYESVLLKAYVNATVKKPKAR 480
QY 481 IIFGGTVIGRS 491
Db 481 IIFGGTVIGRS 491

RESULT 7
AAG32219 ID AAG32219 standard; Protein; 754 AA.
XX AC AAG32219;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 38826.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 03-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140595.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 13-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
```

```
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145319.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149302.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 43.4%; Score 1742; DB 21; Length 754;
Best Local Similarity 48.4%; Pred. No. 1.5e-138;
Matches 370; Conservative 118; Mismatches 237; Indels 40; Gaps 16;

QY 11 IFLFCSSSEILQKQTYIVQLHPNSETAKTFASKEDWHLSPFLOEAVLGVVEEPEPSSR 70
DB 18 LFLLLHTTA-----KKTYILRVN-HSDKPSFLTHDWTYSQLNS-----ESS 59

QY 71 LLYSGAIEGPAQLTESPAE-ILKSPVAVRDPHVLOVQTYISYKFLGLDG-FCNS 128
DB 60 LLXTYTSFHGFSAYLDTSEADSLSSNSILDIPEPLTYLTHTRTEPEFLGNSFVG 119

QY 129 GWSKSRFGOGTIIGVLDTCGWPESSPFDGTGMPSPRKWKGCQGESFSSSCNRKLI 188
DB 120 DLGSSS---NGVIIGVLDTCGWPESSPFDGTGMPSPRKWKGCESGSDPDKLCKKLI 176

QY 189 GARFFIRGHRVANSPEESPNNPREYISARDSTGHTASTVGGSSVSMANVLNGAGVA 248
DB 177 GARSFKGQWASGGGFSK--RESVSPRDVDGHTHTSTTAAGSAVRNASFLGYAAGTA 234

QY 249 RGMAPGAHIAVYKCMFNGCYSSDILAAIDVAIOKVDVLSLSLGGFPPIPLYDDTIAIGT 308
DB 235 RGMATRAVATYKVCNSTGCGFSDILAAAMDRAILDGVDVLSLSLGGSAFYRDTIAIGA 294

QY 309 FRAMERGISVICAAGNNGPIESSVANTAPWVTIGAGTLDRRFPVAVRVLANGKLLYGESL 368
DB 295 FSAMERGVFVSCSAGNSGPTRASVANVAPWMTVGAGTLDRDPFAPANLNGKRLTGVS 354

QY 369 YPGKGIKNAGREVEVIYVTGDKGSEFLRGPREEIRGKVMICDRGVNGRSEKGEAVK 428
DB 355 YSGVGM--GTRPLELVYKNGKSSSSNCLPESLDSISVRGKI VVCDRGVNAVEKGA 412

QY 429 EAGGVAMILANTEINOEDSIDVHLIPATLIGYTESVLLKAYVNATVKPKARIIFGTVI 488
DB 413 DAGOLGMIMANTAASGEELVADSHLLPAAVAGKKTGDLREYVKSDSKPTALLVFKGT 472

QY 489 GRSRAPEVAOFSARGPSLANPSILKPDMIAPGVNIIAAMPQNLGPTGLPYDSRVNFTVM 548
DB 473 DVKPSVVAAFSSRGPNVTPEILKPDVIGVGNILAGWSDAICPTGLDKDSRRTQFNM 532

QY 549 SGTSMSCPHVSGITALIRSAYPNKSAPAAIKSALMTADLYDROGKAIDG--NKPAGVFA 606
DB 533 SGTSMSCPHISGLAGLLKAAHPWSPSAIKSALMTAYVLDNTNAPLHDAADSLNSP 592

QY 607 IGAGHVNPQKAINPGLVYNIQPVDIYTLCTLGTFRSDILAI THK-NVSCNGILRNKPGF 665
DB 593 HSGHVDPOKALSPGLVYDIDISTEYIRFLCSLYTYVDHIVAI VKRPSVNCSSKXF-SDPG- 650
```


PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134421.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134368.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148694.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157885.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144684.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161992.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.
PR 04-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		

Query Match 41.9%; Score 1684; DB 21; Length 775;
Best Local Similarity 44.8%; Pred. No. 1.3e-133;
Matches 354; Conservative 133; Mismatches 260; Indels 44; Gaps 15;


```
Db      527 ASDSRVEFNIIISGTSNCPHVSGLAALLKSVHPWSPAAIRSAALMTTAYKTYKDGKPIIL 586
QY      597 D--GNKPAGVFAAGAGVNPQKAINPGLAVNIQPDVYITLCTGFTSRSDILAIHKVVS 654
Db      587 DIATGSPFPDRGAGHVSPTTAINPGLIYDLTTEDYLGFLCALNYSPOIRSVSRNRYT 646
QY      655 CNGILRNKPNQFS---LNYPISIAVIFKRGKXTTEMITRRVTNVGSPNSIYSVNVKA-PEGIK 710
Db      647 CD----PSKSYSVADLNYSFAVNV-D-GVCAKYKTEVTTSVGAGCT-YSVKYTSSETTGVK 700
QY      711 VIYNPKRLVFKHVDQTLTSYRVWFLKKKNGRGKVASFAOGLTWNHNLMMORVSPISV 770
Db      701 ISVEPAVLNPKANEKKSYTVTFTVDSKPSG---SNSFGSIWSDGKHV---VGSVAI 754
QY      771 T 771
Db      755 S 755

RESULT 12
AAG43556
ID AAG43556 standard; Protein; 780 AA.
XX
AC AAG43556;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 54455.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
```

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

Query Match 40.9%; Score 1642.5; DB 21; Length 780;
Best Local Similarity 43.6%; Pred. No. 4.4e-130;
Matches 342; Conservative 128; Mismatches 281; Indels 33; Gaps 11;

QY	4	KPFFLCI----	IFLFCSSSEILKQTYIVOLHPNSETATFKASKDWHLSFLOEAVLG	59
Dd	9	KPFLFIILSINLIPLQAETTTQISTKTYVLM-DKSAMPLPYTNHLQWSSKINSVTQH	67	
QY	60	VVEEEEPSSRLLYSGSAIEGFAAQLTSEAEILRYSPVAVRPHVLQVQTYSVKF	119	
Dd	68	KSQEEEGNNRIITYTAFFGLAAQLTQEBAELEEDGVVAVIPETRVELHRTSRPTF	127	
QY	120	LGLDGFNGSVMSKRFQGGTHIGVLDTGVMPEPSFDTCMPISIPRKWKICQEGSFSS	179	
Dd	128	LGLERQESRWAEVRTVDHDVVVGVDITGIWFPESESNDTCMSVPATWRACETGRKL	187	
QY	180	SSSNRKLIGARFFIRHRVANSPEESPMPREYISARDSTGHCHTASTVGGSSVSMA	239	
Dd	188	KENCNKIVGARVFYRGVEAATGKIDE--ELEYKSPRRDRDHGTHTAATAVAGSPVKAN	244	
QY	240	VLGNAGVARGMAPGAHIAVYKVCWFNCYCSDILAIDVAIQKDVLDSLGLGPFLP	299	
Dd	245	LFGPAYGTARGMAQKARVAAYKVCMVGCFSDSLISAQDAVDAGVQVLSLSCGGVSTY	304	
QY	300	YDDTIAIGTFRAMERGIVICAAGNNGPIESSVANTAFWSTICAGTLDRFPAPVRLAN	359	
Dd	305	SRDSLSIATFGAMENGVFVCSAGNGGPDPLSLTNVSWITTVASMTMDRDPFPAVIGT	364	
QY	360	GKLLYGESLYPGKIKNAGREVEVIYV---TGCGKGSEFFCLRGSLPREEIRGMVICDRG	416	
Dd	365	MRTFKGVSLYKERTVLPKNQYFLVYLGRNASSPDPTSFCLDGALDRHVAGKLVICDRG	424	
QY	417	VNGRSEKEAVKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYNATVK	476	
Dd	425	VTPRVQKGQVVRKAGGIGNVLNTATNGEELVADSHMLPAVAVGEKEGKLIKQYAMTSKK	484	
QY	477	PKARIIFGCVTLGRSRAPEVAQFSARGPSLANPSLTLPDMTAPGVNIITAAMPQNLGPTGL	536	
Dd	485	ATASUEILTRIGIKPSPVVAAFSSRGPNFUSLEILKPDLAPGVNIILAAWTGDNPSSL	544	
QY	537	PYDSRRVNFTVMSGTSMSCPHVSGITALIRSAYPNWSPPAAIKSALMTTADLYDRQGKAIK	596	
Dd	545	SSDPRVKENILSGTSMSCPHVSGVVAALIKSRHPDWSPAALKSALMTTAYVHDNMFKPLT	604	
QY	597	D--GNKPAGVFAIGAHNVPOKAINPGLVINIQPVDTYITVCTLGFTGTSIDLAIT-HKNV	653	
Dd	605	DASGAAPSPYDHGAGHIDPLRADPGLVYDIGPQEPFEFLCTQDLSPSQLKFVTKHSNR	664	
QY	654	SCNGILRNKPGPSLNYPSTIAVIFKXGKTTEMIT--RRVTNVSPTSIIYVNVKAPGKIKV	711	
Dd	665	TCKHTLAKNPG-NLNYPAISALFPENTHYKAMTLRTVTNVGPHITSYKVSVPFKGASV	723	
QY	712	IYNPKELVFKHVDQTLISY----RVWFVLKKNRGKGKVASFAQGQQTWNVNSHNLMORVRS	767	
Dd	724	TVQPQTLNFTSKHQKLSYTVTFRFRMKRPFEFG-----LWMKST---THKVRSP	771	
QY	768	ISVT	771	
Dd	772	VIII	775	

RESULT 13
ABB93785
ID ABB93785 standard; Protein; 780 AA.
XX ABB93785;
XX DT 31-MAY-2002 (first entry)
XX

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147433.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

Query Match 40.9%; Score 1642.5; DB 21; Length 780;
Best Local Similarity 43.6%; Pred. No. 4.4e-130;
Matches 342; Conservative 128; Mismatches 281; Indels 33; Gaps 11;

QY	4	KPFFLCI----	IFLFCSSSEILKQTYIVOLHPNSETATFKASKDWHLSFLOEAVLG	59
Dd	9	KPFLFIILSINLIPLQAETTTQISTKTYVLM-DKSAMPLPYTNHLQWSSKINSVTQH	67	
QY	60	VVEEEEPSSRLLYSGSAIEGFAAQLTSEAEILRYSPVAVRPHVLQVQTYSVKF	119	
Dd	68	KSQEEEGNNRIITYTAFFGLAAQLTQBEAELEEDGVVAVIPETRVELHRTSRPTF	127	
QY	120	LGLDFGNSGVMSRSRFGOGTHIIGVLDTGVMPEPSFDTCMPISIPRKWKICQEGSFS	179	
Dd	128	LGLERQESRWAEVRTVDHDVVVGVDITGIWFPESESNDTCMSVPATWRACETGRKL	187	
QY	180	SSSNRKLIGARFFIRHRVANSPEESPMPREYISARDSTGHCHTASTVGGSSVSMA	239	
Dd	188	KENCNKIVGARVFYRGVEAATGKIDE--ELEYKSPRRDRDHGTHTAATAVAGSPVKAN	244	
QY	240	VLNGAGVARMAPGAHIAVYKWFNCYCSDILAIDVAIQKDYLVSLSLGGPIPL	299	
Dd	245	LFGPAYGTARGMAQRVAAYKVCMVGCFSDSLISAQDAVDAGVQVLSISLGGVSTY	304	
QY	300	YDDTIAIGTFRAMERGIVICAAGNNGPIESSVANTAFWSTICAGTLDRFPAPVLAN	359	
Dd	305	SRDSLSIATFGAMENGVFVSCAGNGGPDPLSLTNVSWITTVASMTMDRDPFPAVIGT	364	
QY	360	GKLLYGESLYPGKIKNAGREVEVIYV---TGCGKGSEFFCLRGSLPREEIRGNVICDRG	416	
Dd	365	MRTFKGVSLYKERTVLPKNQYFLVYLRGNASSPDPTSFCLDGALDRHVAGKLVICDRG	424	
QY	417	VNGRSEKEAVKEAGGVAMILANTEINOEDSIDVHLLPATLIGTESVLLKAYNATVK	476	
Dd	425	VTPRVQKGQVVRKAGGIGNVLNTATNGEELVADSHMLPAVAVGEKEGKLIKQYAMTSKK	484	
QY	477	PKARIIFGTVIGRSRAPVQAQSARGPSLANPSLTLPDMTAPGVNIITAAMPQNLGPTGL	536	
Dd	485	ATASEILTRIGIKPSPVAAFSRGPENFULELKPDLLAPGVNIILAAWTGDNPSSL	544	
QY	537	PYDSRRVNFTVMSGTSMSCPHVSGITALIRSAYPNWSPAALKSALMTTADLYDROGKAIK	596	
Dd	545	SSDPRVKENILSGTSMSCPHVSGVAAIKSRHPDWSPAALKSALMTTAYVDNMFKPLT	604	
QY	597	D--GNKPAGVFAIGAHVNPQAINPGLVINIQPVDTYITVCTLGFTGTSIDLATF-HKNV	653	
Dd	605	DASGAAPSPYDHGAGHIDFLRATDPLGYDIGPOEYFEFLCTQDLSPSQLKFVTKHSNR	664	
QY	654	SCNGILRNKPGSLNYPSTIAVIFKXGKTTEMIT--RRVTNVGPSNISIYNVAKAPEGIVK	711	
Dd	665	TCKHTLAKNPG-NLNPYPAISALFPENTHYKAMTLRTVTNVGPHISSYKVSVPFKGASV	723	
QY	712	IYNPKELVFKHVDOTLSY----RVWFVLKKNRGKGVASFAGQGQTLWNVSHNLMOVRSP	767	
Dd	724	TVQPCTLNFTSKHQKLSYTVTFRFRMKRPFEFG-----LWKST---THKVRSP	771	
QY	768	ISVT 771		
Dd	772	VIIIT 775		

RESULT 13
ABB93785
ID ABB93785 standard; Protein; 780 AA.
XX ABB93785;
XX DT 31-MAY-2002 (first entry)
XX

DE XX Herbicidally active polypeptide SEQ ID NO 2996.
KW XX Herbicidal; plant; agriculture; herbicide.
XX XX
OS XX Arabidopsis thaliana.
XX XX WO200210210-A2.
XX XX 07-FEB-2002.
XX XX 28-AUG-2001; 2001WO-EP09892.
XX XX 28-AUG-2001; 2001WO-EP09892.
XX XX (FARB) BAYER AG.
XX XX Tietjen K, Weidler M;
XX XX WPI; 2002-269010/31.
XX XX Identifying plant target proteins for herbicidally active compounds,
XX XX comprising aligning and comparing nucleic acid or amino acid sequences
XX XX from plant with nucleic acid or amino acid sequences from non-plant
XX XX organisms -
XX XX
XX XX Claim 5; SEQ ID NO 2996; 261pp + Sequence Listing; English.
XX XX
XX XX The invention relates to identifying target proteins
XX XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX XX aligning and comparing nucleic acid or amino acid sequences from plant
XX XX with nucleic acid or amino acid sequences from non-plant organisms using
XX XX suitable search parameters, where plant sequences having an E-value
XX XX greater by a factor of 3 than the E-value of most similar non-plant
XX XX sequences are selected. The polypeptides or nucleic acids encoding them
XX XX are useful for identifying modulators. The identified modulators are
XX XX useful as herbicides.
XX XX
XX XX Sequence 780 AA;
XX XX
XX XX Query Match 40.9%; Score 1642.5; DB 23; Length 780;
XX XX Best Local Similarity 43.6%; Pred. No. 4.4e-130;
XX XX Matches 342; Conservative 128; Mismatches 281; Indels 33; Gaps 11;
XX XX
QY 4 KPFLCI-----IFLFCSSSEILQQTIVVQLHNSSETAKTPASKFDWHLFLQBAVLG 59
DB 9 KPFLFIILSLINFLQAEITTTQISTKTKYVIHM-DKSAMPLPYTNHLQWSSKINSVTQH 67
QY 60 VEREEEPSSRLYSYGSAEGFAAQLTSEAEILAYSPEVAVRDPHVLQVQTTYSYKF 119
DB 68 KSOEEGNNRILYTYTAFHGLAAQLTQEEARLEEDGVAIVETRYELHTRSPFF 127
QY 120 LGLDGFNSGVWKSREFGQGTIIIGVLDTGVPSPSFDGTGMPISIPKWKGCQGESEFS 179
DB 128 LGLERQSERVWAERVTDHVVGVLDTGVPSESEFNDTGMSFPVATWRGACETGRFL 187
QY 180 SSCNRKLIIGARFIRHVRVANSPESPNMPREYISARSTGHTHTASTVGGSSVSMAN 239
DB 188 KRCNCRKIVGARFVRYGAEATGKIDE---ELEVKSPDRDGHCTHTAATVAGSPYKGAN 244
QY 240 VLNGAGVARGMAPGAHIAKYKWCWNGCYSSDILAAIDVAIQDKVDVLSLGGPPIPL 299
DB 245 LFGFAYTAGMAQAKARVAAYKYKWCWGGCFSSDILSAVDQAQADVQVLSISLGGVSTY 304
QY 300 YDDTIAIGTFRAMEGISVICAAGNNGPIESSVANTAPVSTIGACTLDRRPPAVURLAN 359
DB 305 SRGLSATATGAMEMGVFVSCSAGNGGPDISTNVSPMITTVGASTMDRDPFATVIGT 364
QY 360 GKLYGESLYPGKGIKNAGREVEIYV---TGDKGSEFCRLGSLPREIRGKMWICDRG 416
DB 365 MRIFKGVSLYKGRVLPKQYPLVYIGRNASSPDPTSCFLDGLDRHVGAKIVICDRG 424
QY 417 VNGRSEKGEAVKAGGVAMTILANTEINQEEEDSDVHLLPATLIGYTESVLLKAYNAVTK 476
DB 417 VNGRSEKGEAVKAGGVAMTILANTEINQEEEDSDVHLLPATLIGYTESVLLKAYNAVTK 476

DB 425 VTPRVQKQVVRKAGGIGMVLNTATNGEELVADSHMLPAVAVGEGEKGKLIKQYAMTSKK 484
QY 477 PKARIIFGGTVIGSRRAPEVAQFSARGPSLANPSILKPDMLAPGVNIIIAAPQNLGPTGL 536
DB 485 ATASLEILGTRIGIKPSVVAASFSSRGPNFLSLILKPDMLAPGVNIIIAAPQNLGPTGL 544
QY 537 PYDSRRVNFVTWSGTSMSCPHVSGITALLIRGAYPNWSPAAIKSALMTTADLYDRQGAIK 596
DB 545 SSDPRVKFNILSGTSMSCPHVSGVAALIKSPHDPSPAAIKSALMTTAYVVDNMFELT 604
QY 597 D--GNKPAGVFAIGAGHVNPOKAINPGLVNIQVYITYCTLGFTSRDILAIT-HQNV 653
DB 605 DASGAAPSSPYDHGAGHIDPLRATDPLVYDIGPQEFYFELCTQDLSPLSKVETKISNR 664
QY 654 SCNGILRNKPGFSLNYPISIAVIFKRGKTEMIT--RRVTNUGSPNSISYNNKAPGKIV 711
DB 665 TCKITLAKNPG-NLNYPAISALFPENTHVKAMTLRTVTNYPHISYKSVSPFKGASV 723
QY 712 IVPKRLVFKHVDOTLSY---RVVFLVKKKRGKVASFAQGOLTWVNSHMLQVRVSP 767
DB 724 TVQPKTLNFTSKHQKLSYTVTRFRMKRPEFG-----LVWKST---THKVRSP 771
QY 768 ISVT 771
DB 772 VIIT 775

RESULT 14
AAG43558
ID AAG43558 standard; Protein; 735 AA.
XX AC AAG43558;
XX AC AAG43558;
DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 54457.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
OS Arabidopsis thaliana.
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126284.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0160989.

PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	40.1%;	Score 1612;	DB 21;	Length 735;
Best Local Similarity	44.7%;	Pred. No. 1.6e-127;		
Matches 332;	Conservative 119;	Mismatches 263;	Indels 28;	Gaps

Qy	42	FASKFDWHL\$FLOEAVLGVEEBEPSSRLYSYGSALGFAAQLTESAEILRYSPEV	101
Db	5	YTHNLOWYSKKNSTVQHKSDEEGNNRIITYTQTAPHGLAAQLQEAEERLEEDGV	64
Qy	102	AVRPDRHVLQVTTSYKFLGDFGNCGVNSKSRFGQGTIIGVLDTGVMPEPSFDDTCM	161
Db	65	AVIPETRYELHTTRSPFTFLGERQESERVWAERTVDHDVVVGVLDTGINPESFNDCM	124
Qy	162	PSIPRWKKICQGESFSSSCNKLIGARFFIRGHVRANSPESPNNPREYISARDSTG	221
Db	125	SPVPATWRGACGTGRFLKRCNKRIKGARFYFRGYEAATGKIDE--ELEYKSPRRDGG	181
Qy	222	HGHTASTTVGGSSSVAMVNLNGAGVARGMAPGAHIAVKVCWFNGCYSSDILAADVAI	281
Db	182	HGHTHAATVAGSPVKGALNFPAVGTARGMAQAKARVAAYKVWVGCFSSDILSAVDQAV	241
Qy	282	QDKVDVLSLSLGFFPIPLYDDTTAIGTFPRAMERGISIVICAAGNNGPIESSVANTAPWST	341
Db	242	ADGVQVLSISLGGVSTVSRSLSLATPGAMEMGVFVSCSAGNGGPDPLSLTNVSPWIT	301
Qy	342	IGAGTLDRFPAPVIRLANGLLYGESLYPGHKIKNAOREVEVIYV---TGGDKGSEFCUR	398
Db	302	VGASTMDRDPFATVKIGMTFKGVSLYKGETVLPKNKQYPLVYLGNASSPDP7SFCLD	361
Qy	399	GSI.PREEIRGKMVICDRGVNRSEKGEAVKEAGVAMLANTEINOEDSIDVHLLPATL	458
Db	362	GALDRHVAGKI VICRGVTPRVQKGVVKKAGIGMWLTNTATNGEELVADSHMLPAVA	421
Qy	459	IGYTESVLLKAYVNATVKPARIIPGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMLA	518
Db	422	VGEKGLIKQYAMTSKKATASLETGLTRIGIKPSPVVAAFSSRGPNFLSELKPDLLA	481
Qy	519	PGVNIIAAMPONLGTGPLPYDSRRNVFTVMGTSMSCPHVSIGTALIRSAYPNWSPAAK	578
Db	482	PGVNIIAAMTGMAPSLSLSDPPRRKYENILSGTSMSCPHVSVAALIKSRHPDWSPAAK	541
Qy	579	SALMTTADIYDRQGKAIXD--CNPAGVFAIGAGHVNPQKAINPGLVNIATQPVDITYLIC	636
Db	542	SALMTTAYVHDNMFXPLDASGAAPSSPYDHGAGHI DPLRATDPLGYVDYIGPOEYFEFC	601
Qy	637	TLCGTRASDILAIT-HKNVSCNGILRNKPFGSLNTPSYIAVIFKRGKTIMIT--RRVTNNG	693
Db	602	TQDLSPSQLKVPTKHSNRTCKHTLAKNPG-NLNPALSALFPENTHVKAWTLERTVTNNG	660
Qy	694	SPNSIYSVNVKAPEGIKVINPKRLVFKHVDQTLISY----RVWFVLKKXVRGSKVASPAQ	749
Db	661	PHISSYKVSVPFKGASVTVPQKTLNFTSKHOKLSYTVTPTFRFMKRCPBEFGG-----	713
Qy	750	GQLTWNSENLMQRVRSPISVT	771
Db	714	--LVWKST---THKYRSPVIIIT	730

RESULT 15
AAG43557
ID AAG43557 standard; Protein; 740 AA.
XX AC
AC AAG43557;

RESULT 15
AAG43557
ID AAG4
XX
AC AAG4

```
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142577.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 40.1%; Score 1612; DB 21; Length 740;
Best Local Similarity 44.7%; Pred. No. 1.6e-127;
Matches 332; Conservative 119; Mismatches 263; Indels 28; Gaps 9;

Qy 42 FASKFDHLSFLOEAVLGVVEEPEEPPSSLLSYSGAIEGFAAQLTSEAEILRYSPFV 101
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 10 YTNHLQWSSKINSVTQHKSQEENNRILYTYQTAHFHGLAAQLTQEEAELEEDGV 69
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 102 AVRPDHLVQVQTTYSYKFLGLDGFNGSVWSKSRFCQGTIIIVLDTGVWPEPSDDTGM 161
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 AVIPETRYELHTTRSPFTFLGERQESERVWAERVTDHVVGVGLDTGIWPESESFNDTGM 129
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 162 PSIPRWKIGICQEGESFSSSSCNKRKLGARFFIRHRVANSPEESNMPREVISARDSTG 221
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 SPVPATWRGACETGKFLKNCNKRKIVGARVYRGYEATGKIDE---ELEYKSPDRUG 186
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 222 HGTHASTYGGSSVSMANVLGNAGVARGVAPCAHIAVYKVCWFNGCYSDILAAIDVAI 281
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 HGTHAATVAGSPVKGANLFGPAYGTARGVAKARVAAYKVCWVGCGCFSSDILSAVDQAV 246
: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 282 QDKVDVLSLGGFPDLYDDTTAIGTFRAMERGISVICAAGNNGPIESSVANTAPWVST 341
: : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db      247 ADGVQVLSISLGGVSTYSRDLSLSIATFGAMEMGVFVCSAGNGGPDPIISUTNVSPWITT 306
QY      342 IGAGTILDRFPFAVRLANGKLYGSLYPKGKIGKNAGREVEIYV---TGGDKGSEFCLR 398
Db      307 VGASTWDRDFPATVKTGTRFKGVSLYKGRIVLPKNQYPLVYLGRNASSDPDTSFCLD 366
QY      399 GSLPREIEIKGKVICDRGVNGRSEKGEAVEKEAGGVAMILANTEINOEDSIDVHLLPATL 458
Db      367 GALDRRHVAGKIVICDRGVTFRVQKGVVVRAGGIGWLTNTATNGEELVADSHMLPAVA 426
QY      459 IGYTESVLKAYNAIVKPKARIIFGGTVIGRSRAPEVAQPSARGPSLANPSILKPDMA 518
Db      427 VGEKEGLIKQYAMTSKKATASLEILGTIGIKPGFPVAAAFSSRGPFSLSEILKPDLLA 486
QY      519 PGVNIIAWPNQGLPTGLPYDERRNVFTVMSGTSMSCPHVSGITALIRSAVFNWSPAAIK 578
Db      487 PGVNIILAAWTGDWAPSSLSDDPRVKFNILSGTSMSCPHVSGVAALIKSRHPDWSPAIK 546
QY      579 SALMTTADLYDRQKAID--GNKPAGVPAIGAGHVNPOKAINPGLVYNIQFVDYITYLC 636
Db      547 SALMTTAYVHDNMFKPLTDASGAAPSSPYDHGAGHIDPLRATDPGLVYDIGPQEVFEFLC 606
QY      637 TLGFTFRSDILAIT-HQNVSCNGILRKNPGFSLNYPISIAVIFKRGKTTMIT--RVTNVG 693
Db      607 TODLSPSQLKVETKHSNRTCKHTLAKNPG-NLNYPAISALFPENTHVKAMTLRRTVTNVG 665
QY      694 SPNSIYSVNVKAPGKIVNPKRLVFKHVDQTLSY---RVWFVLKKKNRGKGVASFAQ 749
Db      666 PHISSYKVSVPFGASVTVPKTNFTSKHOKLSYTTVTRFRFRKRPFGG----- 718
QY      750 GOLTWVNSHNLQVRSPISVT 771
Db      719 --LVWKST---THKVSPIIIT 735
```

Search completed: February 2, 2004, 09:10:01
Job time : 49 secs

US-08-894-818B-5

[illegible]

RESULT 2

```

US-08-894-818B-35
; Sequence 35, Application US/08894818B
; Patent No. 6261822
;
; GENERAL INFORMATION:
;
; APPLICANT: TAKAKURA, Hikaru
;
; APPLICANT: MORISHITA, Mio
;
; APPLICANT: YAMAMOTO, Katsuhiko
;
; APPLICANT: MITTA, Masanori
;
; APPLICANT: ASADA, Kiyozo
;
; APPLICANT: TSUNASAWA, Susumu
;
; APPLICANT: KATO, Ikunoshin
;
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
;
; NUMBER OF SEQUENCES: 42
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Browdy and Neimark
;
; STREET: 419 Seventh Street N.W., Ste. 300
;
; CITY: Washington
;
; STATE: D.C.
;
; COUNTRY: United States of America
;
; ZIP: 20004
;
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk.
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-894-818B-35

```

Query Match      8.3%; Score 332; DB 3; Length 654;
Best local Similarity 22.4%; Pred. No. 4.7e-21;
Matches 140; Conservative 65; Mismatches 147; Indels 274; Gaps 21;

QY      29  IVOLHPNSETAKTFASKFDWHLFSFLQBAVLGVDEEEEPSSRLLYSYGS-----AIEG 81
Db      50  IQKLNFEELSTIV--FENH-----REKEIAVRVLELKGAKVRYVYHIIPA 94

QY      82  FAAQ-----LTESEAEI--LRYSEPVAVR-----PDHVLQVQTYTSYKFL 120
Db      95  IAADLKVRDLLVLSGLTGKKALKSGVRFIQEDYKVTSAELEGLEDSEAAQVMATY----- 149

QY     121  GLDGFNGSGWKSRSRGQGTIIQVLDTGWFPSPSPDDTCMPISIPRKWKGICQEGESFSS 180
Db     150  -----WNINLYGDSGGITIGITGTI-----DASHPOL----- 176

QY     181  SSCNRKLIGARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHTAFTVGGSSVSMANV 240
Db     177  ---QGVIGWDFVNG-----RSY--FYDDHGCHGTHVASIAAGTGAA----- 213

QY     241  LNGAGVARGMAPGAHIAVYKVCWFNGCTYS--SDILAAIDVAIOOK-----VDVLSLSLGGF 235
Db     214  ---SNGKYGMAPGAKLAGIKVLGADGSGSISITIIKGVEWAVDNCKYGIKVINLSLGSS 270

QY     296  PIPLYDDTTAIGTFRAWERGISVICAAGNNGPTSESSVANTAPVSTIGAGTLDRRFPADV 355
Db     271  QSSDGTALSOAVNAAWDAGLVVVAAGSGPNKYTIIGSPAASKUITVGADV----- 324

QY     356  RLANKLLYGESLYPGKGIKNAGREVEVITYVTGGDKGSEFCRLGSLPREEIRGMWVICDR 415
Db     325  ----- 324

QY     416  GVNGRGEKGEAVKEAGGVAMILANTEINQBEDSIDVHLLPATLLGYTESVLLKAYVNATV 475
Db     325  -----YDV----- 327

QY     476  KPKARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAAWPNQLGPTG 535
Db     328  -----ITSFSRGPTA--DGRLKPEVAPGNWIIAARASGTS--WG 364

QY     536  LPYDSRRNPNVTMSGTSMCPHVSIGTALIRSYAPNWSPAAIKSAUMTTADLYDROGKAI 595
Db     365  QPIND---YYTAAPGTSMAHPHVAGIAALLQAPHSWTPDKVKTALIEIADV----- 414

```


QY 525 AAMPQNLGPTGLPYDSRRVNFVMSGTSMSCHVSGITALIRSAVNPNSPAAIKSALMTT 584
 Db 223 AARASGTS-MGQPIN--YYTAAPGTSMTATPHVAGIAALLQAHPSWTPDKVKTALET 278
 QY 585 ADLYDRQGAIKDGNKPAV--FAIGAGHVNPOKAIN 619
 Db 279 ADIV-----KPDEIADIYAGGRVNAVYKAIN 304

RESULT 7
 US-08-894-818B-3
 ; Sequence 3, Application US/08894818B
 ; Patent No. 6261822
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: YAMAMOTO, Katsuhiko
 ; APPLICANT: MITTA, Masanori
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: TSUNASAWA, Susumu
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Browdy and Neimark
 ; STREET: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/894,818B
 ; FILING DATE: 20-MAY-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP96/03253
 ; FILING DATE: 07-NOV-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 323285/1995
 ; FILING DATE: 12-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: TAKAKURA=1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 522 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FEATURE:
 ; OTHER INFORMATION: /note= Xaa at position 428 is Gly or Val.

US-08-894-818B-3
 Query Match 8.1%; Score 324.5; DB 3; Length 522;
 Best Local Similarity 23.0%; Pred. No. 1.5e-20;
 Matches 119; Conservative 51; Mismatches 116; Indels 231; Gaps 15;
 QY 110 QVQTTYSYKFLGLDGFNGSVKSRFGQGTIIIGVLDTGWVPESFDDTGMPSPTRKW 169
 Db 12 QVWATY-----VWNLGYDGGSGITIGIDTGI-----DASHPDL----- 44
 QY 170 GICQGESFSSSSCNKLCARFFIRGHRVANSPEPSNMPREYISARDSTGHGHTAST 229

Query Match 8.1%; Score 324.5; DB 3; Length 522;
 Best Local Similarity 23.0%; Pred. No. 1.5e-20;
 Matches 119; Conservative 51; Mismatches 116; Indels 231; Gaps 15;
 QY 110 QVQTTYSYKFLGLDGFNGSVKSRFGQGTIIIGVLDTGWVPESFDDTGMPSPTRKW 169
 Db 12 QVWATY-----VWNLGYDGGSGITIGIDTGI-----DASHPDL----- 44
 QY 170 GICQGESFSSSSCNKLCARFFIRGHRVANSPEPSNMPREYISARDSTGHGHTAST 229

Db 45 -----QKVIQWVDFVNG-----RSY--PYDDHGHGTHVASI 74
 QY 230 VGGSSVSMANVLGNAGVARGMAPGAHIAVYKVCWFGCYS--SDILAAIDVAIQDK--- 284
 Db 75 AAGTGAA-----SNGKYKGMAPGAKLAGIKVLGADGSGSIITIKGVEWAVDNKDKYG 127
 QY 285 VDVLSLSLGGFPPIPLYDDTIAICTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGA 344
 Db 128 IKVINLSLSSQSSDGTALSOAVNAADAGLVVVAAGNSGPNKTYTIGSPAAASKVITV 187
 QY 345 GTLDRRPPAVRLANGKLLYGESLYPGKIGKKNAGREVEVIYVTGDKGSGFCLRGLSPRE 404
 Db 188 GAVDK----- 192
 QY 405 EIRGKMWICDRGVNRSEKGEAVKEAGVAMILANTEINQEDSIDVHLLPATLIGYTES 464
 Db 193 -----YDV----- 195
 QY 465 VLLKAYVNAVTKPKARIIFGGTVIGRSRAPEVAQFARGPSLANPSILKPDMTAPGNWII 524
 Db 196 -----ITSFSSRGPTA--DGRLLKPEVVAPGNWII 222
 QY 525 AAMPQNLGPTGLPYDSRRVNFVMSGTSMSCHVSGITALIRSAVNPNSPAAIKSALMTT 584
 Db 223 AARASGTS-MGQPIN--YYTAAPGTSMTATPHVAGIAALLQAHPSWTPDKVKTALET 278
 QY 585 ADLYDRQGAIKDGNKPAV--FAIGAGHVNPOKAIN 619
 Db 279 ADIV-----KPDEIADIYAGGRVNAVYKAIN 304

US-09-445-472-4
 ; Sequence 4, Application US/09445472
 ; Patent No. 5358726
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: SHIMOJO, Tomoko
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 ; FILE REFERENCE: TAKAKURA=6
 ; CURRENT APPLICATION NUMBER: US/09/445,472
 ; CURRENT FILING DATE: 1999-12-06
 ; PRIOR APPLICATION NUMBER: 151969/1997
 ; PRIOR FILING DATE: 1997-06-10
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 522
 ; TYPE: PRT
 ; ORGANISM: Pyrococcus furiosus
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (428)..(428)
 ; OTHER INFORMATION: Xaa at position 428 is Gly or Val.
 US-09-445-472-4
 Query Match 8.1%; Score 324.5; DB 4; Length 522;
 Best Local Similarity 23.0%; Pred. No. 1.5e-20;
 Matches 119; Conservative 51; Mismatches 116; Indels 231; Gaps 15;
 QY 110 QVQTTYSYKFLGLDGFNGSVKSRFGQGTIIIGVLDTGWVPESFDDTGMPSPTRKW 169
 Db 12 QVWATY-----VWNLGYDGGSGITIGIDTGI-----DASHPDL----- 44
 QY 170 GICQGESFSSSSCNKLCARFFIRGHRVANSPEPSNMPREYISARDSTGHGHTAST 229
 Db 45 -----QKVIQWVDFVNG-----RSY--PYDDHGHGTHVASI 74
 QY 230 VGGSSVSMANVLGNAGVARGMAPGAHIAVYKVCWFGCYS--SDILAAIDVAIQDK--- 284

Db 75 AAGTGAA-----SNGKYKGNAPGAKLAGIKVLGADGSGSISTIIKGVWAVDNKDKYG 127
Qy 285 VDVLISLGGPPIPLDYDDTIAIGTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGA 344
Db 128 IKVNLGSSOSSDGTALSOQVNAWDAGLVVVAAGNSGPKYTIIGSPAAASKVITV 187
Qy 345 GTLDRFPFAVRLANGKLLYGBESLYPGKIKNAGREVEVIYTGDKGSEFCRLGSLPRE 404
Db 188 GAVDK-----YDV----- 192
Qy 405 EIRGKVICDRGVRSEKGEAVKEAGVAMILANTEINQEEDSIDVHLLPATLIGYTES 464
Db 193 -----YDV----- 195
Qy 465 VLLKAVVNAVTPKARIIFGGTVIGRSRAPEVAQFSARGPSLANPSILKPDWIAPGVNI 524
Db 196 -----ITSFSSRGPTA--DGRUKPEVAVPGVNI 222
Qy 525 AAWPNQLGTPGLYDSRRVNFVTMSGTSMCPHVSIGITALIRSAYPNWSPPAAIKSALMTT 584
Db 223 AARASGTS--MGQIPND---YVTAAPGTSWATPHVAGIAALLLQAHPSWTPDKVKTALLET 278
Qy 585 ADLYDRQKAIKGNKPNAGV--FAIGAGHVNPQKAIN 619
Db 279 ADIV-----KPDEIADIAGACRVNAVYKAIN 304

RESULT 9
US-09-000-016-4
; Sequence 4, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 734 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-000-016-4

Query Match 7.4%; Score 297; DB 3; Length 734;
Best Local Similarity 21.4%; Pred. No. 8.2e-18;
Matches 155; Conservative 70; Mismatches 224; Indels 276; Gaps 22;
Qy 36 SETAKTFAS-KPDWHLSEFQEAFLGVVEEPEPSSRLLYSGSAIEGFAAQ----- 85
Db 99 ADAARLVASGKDLRLFDITELGKAATRNQSQGLKVIIVGYQGAARAAKAEVREAGELR 158
Qy 86 -LSEAEILRYSPEVAVRPHVLQVQTVTS-YKFLGLDG-----FNGSGV 131
Db 159 TLTSLNADAVRPHEDASELWDVINGDRTASGIAHVLWDGVVRAALDTSVQIGAPKAW 218
Qy 132 SKSRFOGGTIIGVLTGVWPSPSPDDTGMPSPKPKWKIGICOEGESFSSSSCNRLKIGAR 191
Db 219 SAGYDCKGVKIAVLDTGV-----DTSHPDL-----KGRVTASKNFTAAP----- 257
Qy 192 FFIRHRVANSPESPMPREVIISARDSTGHCHTHTASTVGGSSVSMANVLNGAGVARGM 251
Db 258 -----GAGDKVGHGTHVASIAAGTG-----AQSKGYKGV 287
Qy 252 APGAHTAVKVCWFNSC-YSSDILAAIDVAIQKVDVLSLSLGGFPPIPLYDDTIAIGTFR 310
Db 288 APGAALNKGVLDDSGFGDDSGILAGMEWAAAGQADVVNMSLGGMDTETDPLEAAVKL 347
Qy 311 AMERGIVTCAAGNNGPIESSVANTAPWVSTTICAGTLDRRFPVAVVRLANGKLLYGESLYP 370
Db 348 SAEKGVLFALAGNEGPEISGSPGSADAALTVGA----- 381
Qy 371 GKGIKNAGREVEVIYTGDKGSEFCRLGSLPREIRGKMWICDRGVNRSKGEAVKEA 430
Db 382 -----VDDKDK----- 387
Qy 431 GGVMILANTEINQEEDSIDVHLLPATLIGYTESVLLKAVVNAVTPKARIIFGGTVIGR 490
Db 388 ----- 387
Qy 491 SRAPEVAQFSARGPSLANPSILKPDWIAPGVNIIAWPO-----NLGPTGLPYDSRR 542
Db 388 -----LADFSSTGPRIGDCAI-KPDTAFGVDITASAEGNDIGQEVGSGPAG----- 434
Qy 543 VNFVTMSGTSMCPHVSIGITALIRSAYPNWSPPAAIKSALMTTADLYDRGKAIKGNKPA 602
Db 435 --YMTISGTSMATPHVAGAAALLKQHPDWTSAELKGA-----TGSTKG 477
Qy 603 G---VFAIGAGHVNPQKAINPGLV-----YNIQ-----PV-DIYTLCTLGTFRS 643
Db 478 GKYPTEQSGRIGADKALQQTVIADPVSVSGVQWQPHDDEPVTKQTYR-NLG-TQD 535
Qy 644 DILAITHKNVSCNGILRKPNPGLNYPSTAVIFKRGKTTETLRRVTVNVSNSIYSVNV 703
Db 536 VTUKLTSTATDPKGAAPAGFTLGAITVTVVAGGSASVDM-TADTRLGGTVDGAYSATV 594
Qy 704 KAPEG 708
Db 595 VATGG 599
RESULT 10
US-09-514-340-4
; Sequence 4, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIDINE D
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.

;; COUNTRY: U.S.A.
;; ZIP: 20006
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: Wordperfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/514,340
;; FILING DATE: 28-Feb-2000
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/000,016
;; FILING DATE: January 30, 1998
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Lee Cheng
;; REGISTRATION NUMBER: 40,949
;; REFERENCE/DOCKET NUMBER: <Unknown>
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-721-8200
;; TELEFAX: 202-721-8250
;; TELEX: <Unknown>
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 734 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-514-340-4

Query Match 7.4%; Score 297; DB 4; Length 734;
Best Local Similarity 21.4%; Pred. No. 8.2e-18;
Matches 155; Conservative 70; Mismatches 224; Indels 276; Gaps 22;

Qy 36 SETAKTPAS-KFDWHLISFLOEAVLGVVEEESRLLSYGSAIEGFAAQ----- 85
Db 99 ADAARLVASGKLDRLRFDITELGKAATRSQKGLKVIYGYQGAARAAKAEVREAGELR 158

Qy 86 -LTESEAEILRYSPVAVRPHVLQVQTTYS-YKFLGLDG-----FGNSGVW 131
Db 159 TLTSLNADAVRTHEDASELWDVINGDRGTASGIAHVLGDVRRALDTSVQIGAPKAW 218

Qy 132 SKSRFGGTTIGVLDGVPESFSDDTGMPSPTRKWKICQEGESFSSSSCNKRLIGAR 191
Db 219 SAGYDGKGVKVIALDVG-----DTSHPDL-----KCRVTSKNFTAAP----- 257

Qy 192 FFRGHRVANSPEESPMPREYISARDSTGHGTHTASTVGGSSVSMANVLGNGAGVARGM 251
Db 258 -----GAGDKVGHGTHVASIAGTG-----AOSKGYKEV 287

Qy 252 APGAHIAYKVCWFNGC-YSSDILAAIDVAIQKVDVLSLSLGGFFPIPLYDDTIAIGTFR 310
Db 288 APGAAILNGKVLDSGDDSGILAGMEWAAAGADVNMSLGGMDTPTDPLEAAVDKL 347

Qy 311 AMERGIVICAGNNGPIESSVANTAPWSTTGAGTLDRFPVAVRLANGKLLYGESLYP 370
Db 348 SAEGKVLFAIAGNEGPESTGSGSADAALTGA----- 381

Qy 371 GKGIKNAGREVEIYVTGDKGSEFCIRGSLFREEIRGKVICDRGVNRSRSEKGEAVKEA 430
Db 382 -----VDDKDK----- 387

Qy 431 GGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAVVNAVTKPKARIIFGGTVIGR 490
Db 388 ----- 387

Qy 491 SRAPEVAQSFARSGLANPSILKPDMAFGVNIIAAWPC-----NLGPTGLPYDSRR 542
Db 388 -----LADFSTGTGRLGDGAL-KPDVTAPGVDTAASAEGNIDIGQEVGGPGAG----- 434

Qy 543 VNTVMSGTSMSCFHVSGITALIRSAYPNWSAAIKSALMTTADLYDRGKAIDKGNKPA 602

Db 435 --YMTISGTSMATPHVAGAAALLKQHPDWTSAELKGAL-----TGSTKG 477
Qy 603 G---VFAIGAGHVNPKAINPLV-----YNIQ-----PV-DYITVLTCLGTFRS 643
Db 478 KYTFPEQSGRIQADKALQOTVIADPVSVSGVQQWPHDDEPVTQKLTIR-NLG-TQD 535
Qy 644 DILAITHKNVSCNGILRNKPGFSLNYPISIAVIFRGRKTKTEMITRRVTNVGSPNSIYSNVV 703
Db 536 VTLKLTSTATDPKGAAPAGFTLIGATVTVVPAGGSASVDM-TADTRLGGTVDGAYSAYV 594
Qy 704 KAPEG 708
Db 595 VATGG 599

RESULT 11
US-09-000-016-2
; Sequence 2, Application US/090000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERI
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 823 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-000-016-2

Query Match 7.4%; Score 297; DB 3; Length 823;
Best Local Similarity 21.4%; Pred. No. 9.9e-18;
Matches 155; Conservative 70; Mismatches 224; Indels 276; Gaps 22;

Qy 36 SETAKTPAS-KFDWHLISFLOEAVLGVVEEESRLLSYGSAIEGFAAQ----- 85
Db 99 ADAARLVASGKLDRLRFDITELGKAATRSQKGLKVIYGYQGAARAAKAEVREAGELR 158
Qy 86 -LTESEAEILRYSPVAVRPHVLQVQTTYS-YKFLGLDG-----FGNSGVW 131
Db 159 TLTSLNADAVRTHEDASELWDVINGDRGTASGIAHVLGDVRRALDTSVQIGAPKAW 218

Db 595 VATGG 599

RESULT 13

US-09-000-016-7

; Sequence 7, Application US/09000016

; Patent No. 6143541

; GENERAL INFORMATION:

; APPLICANT: Akira ARISAWA et al.

; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC

; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV

; TITLE OF INVENTION: ITS EXPRESSION PRODUCT

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

; STREET: 2033 K Street, N.W., #800

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20006

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

; COMPUTER: IBM Compatible

```

SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.

```

```

NAME: Warren M. Creek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 520 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-000-016-7

Query Match      7.0%; Score 283; DB 3; Length 520;
Best Local Similarity 21.8%; Pred. No. 8.5e-17;
Matches 133; Conservative 56; Mismatches 170; Indels 252; Gaps 18;

QY      126  GNSGWKSKRRPGQGTIIGLVDTGVWPESPSPDDTGMPISIPKWKIGICQEGESFSSSCNR 185
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      9  GAPKWSAGYDGKGVKIAVDTGV-----DTSHPDL-----KGRVTASKNFTAAP--- 53

QY      186  KLIGARFFIRHRVANSPEESPNNPREYISARDSTGHGTHHTASTVGGSSVSMANVLNGA 245
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      54  -----GAGDKVGHGTHVASIAAGTG-----AQS 77

QY      246  GVARGMAPGAHIAIVKVCWFNGC-YSSDILAAIDVAIQDKVDVLVLSLGGFPPIPLYDDTI 304
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      78  GKYGKVPAGGAILNKGVLDDSGFDGDSGILLAGMEWAAQAQGADVYVNMISLGGMDTPETDPLE 137

QY      305  AIGTFRAMERGISVTCAGNNGPIESSVANTAPVWSTIGAGTLDRRFPVAVVLANGKLLY 364
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      138  AAVDKLSAEKGVLPFAIAGNEGPGSISGPGSADAAALTVGA----- 177

QY      365  GESLYPGKGIKNAGREVEVIYVTGGDKSEFCLRGLPREEIRGNKVICDRGVNGRSEK 424
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Q1 178 -----VDPKDK----- 183

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 2, 2004, 09:11:33 ; Search time 38 Seconds
(without alignments)

4238.517 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MEKPPFLCIIFLFCSSSS.....NSHNLQVRISVTLKTN 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1175	29.2	734	12	US-10-259-165-272
2	332	8.3	654	14	Sequence 72, App
3	330	8.2	659	14	Sequence 16, Appl
4	324.5	8.1	412	14	Sequence 12, Appl
5	324.5	8.1	522	14	Sequence 1, Appl
6	318.5	7.9	1208	15	US-10-090-624-4
7	307	7.6	1139	15	US-10-090-624-4
8	282	7.0	1079	15	US-10-156-761-13251
9	276.5	6.9	382	11	US-10-112-488-39
10	271.5	6.8	382	12	US-10-112-488-39
11	271.5	6.8	382	14	US-10-146-905A-8
12	271.5	6.8	382	15	US-10-090-624-31
13	269.5	6.7	1150	10	US-10-104-693-2
14	266.5	6.6	382	12	US-09-870-122-3
15	264.5	6.6	382	11	US-10-146-905A-11
					Sequence 20, Appl

16	262	6.5	1237	12	US-10-314-657-4
17	261.5	6.5	1181	10	US-09-870-122-23
18	260.5	6.5	382	12	US-10-423-649-2
19	260.5	6.5	382	15	US-10-033-325-2
20	260.5	6.5	382	15	US-10-228-572-2
21	259	6.4	418	10	US-09-966-921A-2
22	257.5	6.4	380	11	US-09-813-408-19
23	257.5	6.4	1167	10	US-09-870-122-2
24	255.5	6.4	361	12	US-10-007-389-1
25	251.5	6.3	377	11	US-09-813-408-1
26	251	6.2	380	11	US-09-824-893A-261
27	248	6.2	380	12	US-10-324-152-8
28	248	6.2	380	12	US-10-324-152-9
29	248	6.2	381	10	US-09-920-118-16
30	247.5	6.2	357	9	US-09-837-235-15
31	245	6.1	375	11	US-09-813-408-8
32	244.5	6.1	379	11	US-09-813-408-11
33	243	6.0	269	11	US-09-779-334A-7
34	242.5	6.0	379	10	US-09-920-118-14
35	241	6.0	627	11	US-09-927-827-60
36	240	6.0	275	15	US-10-104-693-3
37	240	6.0	279	11	US-09-813-408-21
38	239	5.9	275	11	US-09-813-408-23
39	239	5.9	275	12	US-10-324-152-1
40	237.5	5.9	274	12	US-10-324-152-1
41	237.5	5.9	379	11	US-09-813-408-10
42	237.5	5.9	379	12	US-10-146-905A-10
43	237.5	5.9	379	15	US-10-202-339-2
44	237	5.9	275	8	US-08-322-678-7
45	237	5.9	275	9	US-09-060-854B-3

ALIGNMENTS

RESULT 1

US-10-259-165-272

; Sequence 272, Application US/10259165

; Publication No. US20030135888A1

; GENERAL INFORMATION:

; APPLICANT: Zhu, Tong

; APPLICANT: Wang, Xun

; APPLICANT: Chang, Hur-song

; APPLICANT: Briggs, Steven P.

; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane

; APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyaki

; APPLICANT: Kreps, Joel

; APPLICANT: Moughamer, Todd

; APPLICANT: Provart, Nicholas

; APPLICANT: Ricks, Darrell

; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING

; FILE REFERENCE: 70030-NP

; CURRENT APPLICATION NUMBER: US/10/259,165

; PRIOR FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/370,620

; PRIOR FILING DATE: 2002-04-04

; PRIOR APPLICATION NUMBER: US 60/368,327

; PRIOR FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26

; NUMBER OF SEQ ID NOS: 782

; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta

; SEQ ID NO 272

; LENGTH: 734

; TYPE: PRT

; ORGANISM: Oryza sativa

US-10-259-165-272

Query Match 29.2%; Score 1175; DB 12; Length 734;

Best Local Similarity 37.1%; Pred. No. 1.1e-102;

Matches 293; Conservative 121; Mismatches 249; Indels 126; Gaps 27;

QY 38 TAKTASKEFDEL-SFLOEAVLGVEEBEERS-----SRLLYSV 75
Db 4 TUSSEPHNVDDWITTSFSLVIVYMGKXDDPESVVTASHHDILTSLVSGKDGAMKSIYVSY 63
QY 76 GSAIBGFAAQLTESAEILRYSFVAVRPHDLVQVOTYSYKFLGLDGFNGSGWKSXR 135
Db 64 KHGFSGFAAMLTESQAEELARLEPVISVKPNTYHQATRSWDFLGLNYSQSLKKA 123
QY 136 FCGITIGVLDGVPWSPSPDDTGMPSIPRKWGI COEGESFSSS--CNKRLIGARPT 194
Db 124 NGEDVIVGVIDSGIWPESRSPDNGYSPVPARWKGKQCTGAFAFNATTCNKKIIGVRWYS 183
QY 195 RGRHVANSPEESPNNPREVISARDSTGCTHTASTVGGSSVSMANVLNG--AGVARGM 251
Db 184 GG-----IPDE--NKGETYSARDJGGHGTHTASTVGGSSVSMANVLNG--AGVARGM 236
QY 252 APGAHIAVYKVCWFNG-----CYSSDILAAIDVAIOKVDVLSLSLGGFPPIPLYDDTTAIG 307
Db 237 APPARVAVYKVCW--GLRAQCGGAALAAIDAMNDGVDVLSLSIGG-----AG 283
QY 308 ----TFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGAGTLDRRPPAVVRLANGKLL 363
Db 284 EHVETLHVARGIPVVGNDGPTPQIVRNTVPWVITVAASTIDRAFFTVISLGNKKF 343
QY 364 YGESLYPGKGIKNAGREVEVIYVGGDKGSEFCRLGSLPREIRCKWICD-----414
Db 344 VQGSLYNATASSTKQMLV-----DGSS--CDTQTLASINITSKVLCSPSPSIMPRL 395
QY 415 --RGVNGRSEKGEAVKEAGGVAMILANTEINOEDSIDV---HLLPATLIGYTESVLLKA 469
Db 396 SIGDIIGR-----VTKAGANGLIFVOYSVSNALDFLNACSRASVPCVLVDYEITRRIES 449
QY 470 YVNATVPKRIIFGCTVIGRS--RAPEVAQFARGPSLANPSILKPDMLAPGVNIIAAMP 528
Db 450 YMTSTSTPMVKYSSAMTVGSGVLSPIAFAFSSRGPSSLPFGILKPDIAAPGVSLAA--507
QY 529 QNLGPTGLPYDSRRVNTVMSGTSMSCPHVSGITLIRSAYPNWSAPKASALMTADLY 588
Db 508 ----VGDSYE-----LKSGETSMACPHVSAVALLKMWHPDMSPAIKSAIVTASVT 555
QY 589 DROGKAID--GNKEPAGVPAFGAGHVPKAINPGLVYNIOVDYITVL--CTLGFTSRD 644
Db 556 DRFGMPIOAEVPRKADPFDGGGHIEBKADPGLVYDIDPSHYTKFTFNCITLPEASDD 615
QY 645 ILAITHKWVSCNGILRKNPFCFSLNPSIAVIFRKGKTEMITRRVNTVNGSPNSIYVNVK 704
Db 616 -----CESYMEQI--YQNLPSIAVPLKDSVT--VNRVTNVGEAEATYHALE 661
QY 705 APEGIKVINPKRLVF--KHVDQTLVYRVFVLKKNRGKGVASFAGQQLTWV--NSHNL 761
Db 662 APVGMTNVEPVSITTRGSRSVTFKVTFTTTQVQGG----YTFGSLTLDGNTHS--715
QY 762 QVRSPISV 770
Db 716 --VRPIAV 722

RESULT 2

US-10-090-624-16
; Sequence 16, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472

RESULT 3

US-10-090-624-12
; Sequence 12, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624

; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; BEST FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-090-624-16

Query Match 8.3%; Score 332; DB 14; Length 654;
Best Local Similarity 22.4%; Pred. No. 3.4e-22;
Matches 140; Conservative 65; Mismatches 147; Indels 274; Gaps 21;

QY 29 IVOLHNSESTAKTFASKFDWHLSPLOEAVLGVEEBEERSRLLYSYGS-----AIEG 81
Db 50 IQKLNPNSEISTVIV--FENH-----REKEIAVRVLELMGAKRVYVHIIIPA 94
QY 82 FAAQ-----LTSEAEI--LRYSPEVAVR-----PDHVLQVOTTYSYKFL 120
Db 95 IADLKVRDLVLSLGTGKAKLSGVRFIOEDYKVTVSAAEGLDESAAQVMATY----149
QY 121 GLDGFNGSGWKSRRFGQGTIIGVLDGTVGWPSPSPDDTGMPSIPRKWGI COEGESFSS 180
Db 150 ----VWNLGYDGSGITIGIITGTI-----DASHFDL-----176
QY 181 SSCNRKLI GARPFIRGHRVANSPEESPNNPREVISARDSTGCTHTASTVGGSSVSMANV 240
Db 177 ---CGKVGIVGVDVFN-----RSY--PYDDHGHGTHVASIAAGTGAA--213
QY 241 LKNGAGVARGMAPGAHIAVYKVCWFNGCYS--SDILAAIDVAIOK---VDVLSLSLGGF 295
Db 214 ---SNKYKGMAPEAKLAGIKVLGADGSGSISTIIKGVWAVDNKDKYIKVNLISLSS 270
QY 296 PIPLYDDTTAIGTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGAGTLDRRPPAVV 355
Db 271 QSSDGTALDQAVNAANDAGLVVVVAAAGNSGPNKYITIGSPAAASKVITVGAVDK----324
QY 356 RLANGKLYGESLYPGKGIKNAGREVEVIYVGGDKGSEFCRLGSLPREIRCKWICDR 415
Db 325 -----YDV-----324
QY 416 GVNRESEKGEAVKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYNATV 475
Db 325 -----YDV-----327
QY 476 KPKARIIFGCTVIGRSRAPEVAQFARGPSLANPSILKPDMLAPGVNIIAAMPQNLGPTG 535
Db 328 -----ITSFSSRGPTA--DGRLKPEVAVPAGNWIILAAASGTS--MG 364
QY 536 LPYDSRRVNTVMSGTSMSCPHVSGITLIRSAYPNWSAPKASALMTADLYDROGKAI 595
Db 365 QPIND---VYTAAPGTSMATPHVAGIAULLQAHPSMTDPKVTALLETADIV-----414
QY 596 KQGNKPAV--PAIGAGHVPKAIN 619
Db 415 ----KPDEIADIAYAGRVNAYKAIN 436

```
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Thermococcus celer
; US-10-090-624-12

Query Match      8.2%; Score 330; DB 14; Length 659;
Best Local Similarity 23.7%; Pred. No. 5.3e-22;
Matches 124; Conservative 57; Mismatches 119; Indels 224; Gaps 17;

QY 103 VRPHVLQVQTTYSYKFLGLDGFNGSWGSKSRFP-CQGIIIGVLDTGVWPSPSPEDDTGCM 161
Db 128 IQEDYKQVDDATSVSQIGAD-----TVNLSGLYDGGVVAIVDTGI-----DANH 174
QY 162 PSIPRWKGIQCGESFSSSSCNKRLIGARFFIRGHRVANSPEESPNPREVISARDSTG 221
Db 175 PDL-----KG-----KVIGWYDAVNGRS-----TPYDDQG 199
QY 222 HGHTASTVGGG-SVSMANVLNGAGVARGMAPGAHIAVYKVCWFNGCYS-SDILAAIDV 279
Db 200 HGTHVAGIVAGTGSVNSQYI-----GVAFGAKLVGVKVLGADGSGSVSTIAGVDW 250
QY 280 AIQDK-----VDVLSLSLGFPPIPLYDDTTAIGTFRAMERGISVICAGNNGPIESSVANT 335
Db 251 VQNKDKYGRVIRNLSLSSQSDGTDLSQAVNNAWDAGLVVVAAGNSGPNVTYVGSF 310
QY 336 APWVSTIGACTLDRPFPVAVRLANGKLLYGESLYPGKIKNAGREVEIYVVTGGDKGSEF 395
Db 311 AAASKVITVGAVDN----- 325
QY 396 CLRGSLPREIRGNWICDRGVNGRSEKGEAVKEAGGVAMILANTEINOEDSIDVHLLP 455
Db 326 ----- 325
QY 456 ATLIGYTESVLLKAYVNATVKPKARIIFGCTVIGRSRAPEVAQFSARGPSLANPSILKPD 515
Db 326 -----DNIAFSRSGPTA--DGRLEKPE 345
QY 516 MIAPGVNIIAAMPQNLGPT-GLPYDSRRVNTVMSGTSMSCPHVSGITALIRSAYPNWS 574
Db 346 VWAPGVDDIIA--PRASGTSMTGTEIND---YYTKASGTSMTATPHVSGVAGLILQAHPSWTP 400
QY 575 AAKSALMTTADYDRQKAIKDGKPKAGV-FAIGAGHVNPQKAI 618
Db 401 DRVKTALITADIV--APKEIAD-----IAYGAGRVNYYKAI 435

RESULT 4
US-10-090-624-1
; Sequence 1, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; FEATURE:

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; US-10-090-624-1

Query Match      8.1%; Score 324.5; DB 14; Length 412;
Best Local Similarity 23.0%; Pred. No. 7.9e-22;
Matches 119; Conservative 51; Mismatches 116; Indels 231; Gaps 15;

QY 110 QVQTTYSYKFLGLDGFNGSWGSKSRFGQGTIIGVLDTGVWPSPSPEDDTGMSIPKWK 169
Db 12 QVWATY-----VMNLGYDGGTIGTIDTGI-----DASHPD 44
QY 170 GICQGEFSFSSSSCNKRLIGARFFIRGHRVANSPEESPNPREVISARDSTGHTHTAST 229
Db 45 -----QGVKIVGWDFVNG-----RSY--PYDDHGHGTHVASI 74
QY 230 VGGSSVSMANVLNGAGVARGMAPGAHIAVYKVCWFNGCYS-SDILAAIDVAIQDK----- 284
Db 75 AAGTGAA-----SNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVVEWAVDNKDKYG 127
QY 285 VDVLSLSLGFPPIPLYDDTTAIGTFRAMERGISVICAGNNGPIESSVANTAPWVSTIGA 344
Db 128 IKVINLSLSSQSDGTDLSQAVNNAWDAGLVVVAAGNSGPNKYTIGSPAAASKVITV 187
QY 345 GTLDRRPFPVAVRLANGKLLYGESLYPGKIKNAGREVEIYVVTGGDKGSEFCLRGSLPRE 404
Db 188 GAVDK----- 192
QY 405 EIRGMWICDRGVNGRSEKGEAVKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTES 464
Db 193 -----YDV----- 195
QY 465 VLLKAYVNATVKPKARIIFGCTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNII 524
Db 196 -----ITSFSSRGPTA--DGRLEKPEWVAPGNWII 222
QY 525 AAMPQNLGPTCLPYDSRRVNTVMSGTSMSCPHVSGITALIRSAYPNWSAAIKSALMTT 584
Db 223 AARASGTS-MQGPIND---YYTAPGTSMTATPHVAGIAALLLQAHPSWTPDKVTALLET 278
QY 585 ADLYDRQKAIKDGKPKAGV-FAIGAGHVNPQKAI 619
Db 279 ADIV-----KPEIADIAYGAGRVNYYKAI 304

RESULT 5
US-10-090-624-4
; Sequence 4, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: (428)..(428)
; OTHER INFORMATION: Xaa at position 428 is Gly or Val.
US-10-090-624-4

Query Match      8.1%; Score 324.5; DB 14; Length 522;
Best Local Similarity 23.0%; Pred. No. 1.2e-21;
Matches 119; Conservative 51; Mismatches 116; Indels 231; Gaps 15;

QY 110 QVOTYSYKFLGDLGFGNSGWSKSRFGGTTIGVLDTGVPSPSPDDTGMPSIPRKWK 169
Db 12 QVMATY-----VMNLGYDGGTITIGIITGTI-----DASHPDL----- 44
QY 170 GIQCEGESFSSSCNRKLIGARFFIRGHRVANSPEESPNNPREYISARDSTGHGHTAST 229
Db 45 -----QGVKIVGVDFVNG-----RSY--PYDDHGHGHTVASI 74
QY 230 VGGSSVSMANVLNGAGVARGMAPGAHIAVYKVCWFGNCGYS--SDILAAIDVAIQK----- 284
Db 75 AAGTGAA-----SNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVEMAVDNKDKYG 127
QY 285 VDVLISLSGFFPLVDYDTTIAITFRAMERGISVICAAGNNGPIESSVANTAPWYSTICA 344
Db 128 IKVINLSLSSQSSDGTDLASQAVNAADAGLVVVVVAAGNSGPNKYTTIGSPAAASKVITV 187
QY 345 GTLDRRFPVAVRLANGKLLYGESLYPGKIKNAGREVEVIYVTGGDKGSEFCIRGSLPRE 404
Db 188 GAVDK-----YDV----- 192
QY 405 EIRGKWCIDRGVGRSEKGEAVKEAGGVAMILANTEINQEBSDVHLLPATLIGYTES 464
Db 193 ----- 195
QY 465 VLLKAYNATVKPKARIIFGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNII 524
Db 196 -----ITSFSSRGPTA--DGRLEPEVYVAPGNVII 222
QY 525 AAMPQNLGPTPLVDSDRRVNTVMSGTSMSCPHVSGITAIIRSAVFNWSPAAIKSALWTT 584
Db 223 AARASGTS--MGQPLND---YTTAAPTGMTATPHVAGVALLQAHPSWTPDKVKYKALLET 278
QY 585 ADLYDRQKKAIDGNKPGV--FAIGAGHVNPOKAIN 619
Db 279 ADIV-----KPEDEIADIYAGRVNAYKAIN 304

RESULT 6
US-10-156-761-13251
; Sequence 13251, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13251
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13251

Query Match      7.9%; Score 318.5; DB 15; Length 1208;
Best Local Similarity 24.5%; Pred. No. 1.9e-20;
Matches 129; Conservative 45; Mismatches 113; Indels 239; Gaps 19;

QY 126 GNSGVWSKSRFGGTTIGVLDTGVPSPSPDDTGMPSIPRKWKIGICQEGESFSSSCNR 185
Db 198 GTRAAWDAGTUGOVTVAVDGTG-----DTHPDL-----AGRVSRKS--- 237
QY 186 KLIGARFFIRGHRVANSPEESPNNPREYISARDSTGHGHTASTVGGSSVSMANVLNGA 245
Db 238 -----FIDBEVA-----DRNGHGHVHTVTVGGSGAA-----SD 266
QY 246 GVARGMAPGAHIAVYKVCWFGNCGYS--SDILAAIDVAIQD--KVDVLSLSLGGF-----P 296
Db 267 GTERGVAPGATLAVGKVLSDQAGAGSQQIAGMEWAARDVRAIRIVMSLSGSTASDGTDP 326
QY 297 IPLYDDTIAITFRAMERGISVICAAGNNGPIESSVANTAPWYSTIGAGTLDLRRFPVAVR 356
Db 327 MAEAVDTL-----SEETGALFVVAAGNTG-----AP--SSTGS----- 357
QY 357 LANGKLLYGESLYPGKIKNAGREVEVIYVTGGDKGSEFCIRGSLPREIRGRKMWICDRG 416
Db 358 -----PGAA----- 361
QY 417 VNGRSEKGEAVKEAGGVAMILANTEINQEBSDVHLLPATLIGYTESVLLKAYNATVK 476
Db 362 -----DS-----ALTVGAVDS----- 372
QY 477 PKARIIFGTVIGRSRAPEVAQFSARGPSLANPSILKPDMIAPGVNIIAAMPQNLGPTGL 536
Db 373 -----SDRAAYFTSAGFRHGD--NALKPDLAAPGVDIRAARSOLAPGTGY 415
QY 537 PYDSRRVNTVMSGTSMSCPHVSGITAIIRSAVFNWSPAAIKSALMTTADLYDRQKKAIK 596
Db 416 -----YTSMTSGTSMATPHVAGVALLAEQHPDWTGARKLDMSTSQLD----- 460
QY 597 DGNKPGVFAIGAGHVNPOKAINPGLVNIQPVYIITLCTLGFTTR 642
Db 461 -----ASYVQLGAGRVSPDVGARVT-----ATGSADLGFHR 493

RESULT 7
US-10-156-761-10856
; Sequence 10856, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10856
; LENGTH: 1139
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10856

Query Match      7.6%; Score 307; DB 15; Length 1139;
Best Local Similarity 23.2%; Pred. No. 2.1e-19;
Matches 153; Conservative 58; Mismatches 132; Indels 256; Gaps 22;

QY 126 GNSGVWSKSRFGGTTIGVLDTGVPSPSPDDTGMPSIPRKWKIGICQEGESFSSSCNR 185
```

Db 241 CAPTAWAGYDGGKAVLDTGDVATHDPLKQ-----VAESKNFSAAA--- 285
Qy 156 KLIGARPPIRHRVANSPEESPNNPREVISARDSTGHGTHASTVGGSSVSMANVLNGA 245
Db 286 -----DAADFEGHGHVASTAAGTG-----AKSN 309
Qy 246 GVARGMAPGAHIAVYKVCWFNGC-YSSDILAAIDVAIQDKVDVLSLSLGGFPPIPLYDDTI 304
Db 310 GYKGVAFGATILNGKVLDDTSGDSDGILAGMEWAEOGADVNLVSLGGDTPEIDPLE 369
Qy 305 AIGTFRAMERGISVICAAGNNGPIESSVANTAPWVSTIGAGTLDRRPPAVVRLANGKLLY 364
Db 370 AENVKLSBEKGILFAIAGNEGE-----F 393
Qy 365 GESLYPGKGIKNAGREVEVIYVTGGDKSEFCLEGSIPREIRGKMWICDRGVNGRSEK 424
Db 394 GE----- 395
Qy 425 EAVKEAGGVAMILANTEINBEDSIDVHLLPATLIGYTESVLLKAYVNVATVKPKARIIFG 484
Db 396 QTIGSPGAADALTGAVNDS----- 417
Qy 485 GTVIGRSAPRAVAFSARGPSLANPSILKPDMIAPGVNIIAA-----WPNLGL--PTGL 536
Db 418 -----KLASFSSRGPL--DGAIKPDVTPAGVDITAAAPGVSVIDQEVQKPDG- 464
Qy 537 PYDSRRVNTVSGTSMSCPHVSGITALLIRSAYNWSPAPAKSALMTTADLYDRQKAIK 596
Db 465 -----YLTISGTSNATPHVAGAAAILKQOHPNWSFAELKALTGSA-----KGGKYTP 512
Qy 597 DGNKPAGVFAIGAHVNPQKAINPGLVNIQPDY-----ITVLCITLGF 640
Db 513 -----FOGSGRIADVKAIKOSVTANPNSVSGTQOHPHTDDKPVTOQLTYR--NLG- 582
Qy 641 TRSDI---LAITHKNVSCNGILRKNPGFSLNYPGSIPIFKEGKTTEMITRRVTNVGSPNS 697
Db 563 -TSDVTLNASTATNP--KGVAAPSGFPKLGATKVTVP--AGGKASVDFTVNTKLGTTD 618
Qy 698 IYVNVKAPGKIVNPKRLVFKHVDQTLRYVNVFVKKNRGKGVASFAQGLTWN 756
Db 619 AYSAVVTATGGQT--VRTAAAVQREVE---SYDV--TLKHIDRDKPANYSTDLTGVS 671

RESULT 8

US-10-112-488-39
; Sequence 39, Application US/10112488
; Publication No. US20030082746A1
; GENERAL INFORMATION:
; APPLICANT: KIKUCHI, Yoshiaki
; APPLICANT: DATE, Masayo
; APPLICANT: UMEZAWA, Yukiko
; APPLICANT: YOKOYAMA, Keiichi
; APPLICANT: MATSUI, Hiroshi
; TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSLUTAMINASE
; FILE REFERENCE: 219286USOCNT
; CURRENT APPLICATION NUMBER: US/10/112,488
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: PCN/JPO0/06780
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: JP2000-280098
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: JP11-280098
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Streptomyces alboborisolus
US-10-112-488-39

Query Match

7.0%; Score 282; DB 15; Length 1079;

Best Local Similarity 20.8%; Pred. No. 4.8e-17;
Matches 158; Conservative 71; Mismatches 257; Indels 272; Gaps 22;
Qy 36 SETAKTFAS-KDMLHSLFLOEAVLQVEEBEESPSSRLIYSYCSAIEGFAAQ----- 85
Db 67 ADARILVASGKLDQRLFDVTELNKAATTAHGGUKVIVGVYGAAGAADVRDAGTVRR 126
Qy 86 -LSESEAIRLYSPVAVRDPHVLQVQTTYS-YKFLGLDG-----FNSGVW 131
Db 127 TLTSLNADAVQTPQAGAELEWAVTDGDTASGVARVWLDGVRKASLDTSVQIGTPKAW 186
Qy 132 SKSRGQGTIIQVLTIGVWPSPSPDDTGMPISIPRKNWGIQEGESFSSSSCNRLKIGAR 191
Db 187 EAGYDGGYKIAVLDTGV-----DATHPDL-----KQVTSKNTFSAP----- 225
Qy 192 FIIRGHRVANSPEESPNNPREVISARDSTGHGTHASTVGGSSVSMANVLNGAGVARGM 251
Db 226 -----TTGDVVGHGTHVASTAAGTG-----RQSKTYTKGV 255
Qy 252 AFGAHIAVYKVCWFNGC-YSSDILAAIDVAIQDKVDVLSLSLGGFPPIPLYDDTIAGTFR 310
Db 256 AFGAKILNGKVLDDAGFGDDSGILAGMEWAAGADIWNLSLGGWDTTETDPLEAAVDKL 315
Qy 311 AMERGISVICAAGNNGPIESSVANTAPWVSTIGAGTLDRRPPAVVRLANGKLLYGESLYP 370
Db 316 SAEKGILFAIAGNEGPQSIGSPGSADSALTGVA----- 349
Qy 371 GKGKINAGREVEVIYVTGGDKSEFCLEGSIPREIRGKMWICDRGVNGRSEKGEAVKEA 430
Db 350 -----VDDKDK----- 355
Qy 431 GGVAMILANTEINBEDSIDVHLLPATLIGYTESVLLKAYVNVATVKPKARIIFGTVIGR 490
Db 356 ----- 355
Qy 491 SRAPEVAQSARGPSLANPSILKPDMIAPGVNIIAAMPQNLGPTGLPYDSRRVNTVMSG 550
Db 356 -----LADFSSTGPRLGAV-KPDLTAPGVDITAAKAG-NDIAKEVGEKPAGYMTISG 408
Qy 551 TSMSCPHVSGITALLIRSAYNWSPAPAKSALMTTADLYDRQKAIKQGNKPAVFAIGAG 610
Db 409 TSNATPHVAGAAAILKQOHPENKAYELKALTAST-----KDGKYTP--FEQSG 456
Qy 611 HYNPKAINPGLVNIQPDYITVLTCLGF-----TRSDILAIATHKNVSCNGILRK 661
Db 457 RVQVDKAITQTVI--AEPV-----SLSFGVQOHPHADDKEVTKKLYRNLTGDEDTLK 507
Qy 662 -----NPGFSLNYPGSIPIFKEGKTTEMITRRVTNVGSPNSIYSVNVKAPEG 708
Db 508 LTSTATGPKGAAPAGFFTLGASTLTVPANGTASVDVTADTDLGAVDGTYSAYVVA-TG 566
Qy 709 IKVINPKRLVFKHVDQTLRYVNVFVKKNRGKGVAS 746
Db 567 AQOSVRTAAAVEREVE---SYNV--TLKVLDRSGKATA 599

RESULT 9

US-09-813-408-7
; Sequence 7, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Marrs, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides
; TITLE OF INVENTION: Of Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 382
; TYPE: PRT

; ORGANISM: Bacillus sp.
US-09-813-408-7

Query Match 6.8%; Score 276.5; DB 11; Length 382;
Best Local Similarity 20.8%; Pred. No. 2.7e-17;
Matches 127; Conservative 62; Mismatches 177; Indels 245; Gaps 21;

```
QY 8 LCIIIFLLFCSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSEFLQEAVLGVVEEBEE 67
DB 14 LLLSILATSVSAEEQKKVLIQFE-NQLQVTEFVSSDKQS---ENSLFAEVNDESI 69
QY 68 SSRLLSYGSAIEGFAAQLTESEAEILRYSPVAVRPHVLQVQTTYSYKFLGLDGF 127
DB 70 EMELLYEF-EDIPVVSVELSPEDVKLEKXPSITYIBED--IEVTITNQVTPGIRVOA 126
QY 128 SGVWSKRFQGGTIIIGVLDTGVPSPSFDGTGMPSPRKKWKGICQEGESFSSSCN 197
DB 127 PTAWTRGYTGVRVAVLDGTG-STHPLNIRG-----GVSF----- 162
QY 188 IGARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHASTVGGSSVSMANVLNGAG 247
DB 153 -----VPGEP-----SYQDNGHGHVAGTIA-----ALNNSIGV 192
QY 248 ARGMAPGAHIAVYKVMFNGCYS-SDILAAIDVAIQKVDVLSLGGFFPIPLYDDTIAI 306
DB 193 V-GVAPNAELVAVKVLGANGSGSVSSIAQGLQNTAQNNIHVANLSLGS---PVGSQ 248
QY 307 GTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTILDRFPFAVVRLANGKL 366
DB 249 AVNQATNAGVLVAATGNN-----GSGTVS----- 273
QY 367 SLYPGKGIKNAGREVEVIYVTGGDKGSEFCRLGSLPREIRGKMWICDRGVNGRSEK 456
DB 274 --YPAR----- 277
QY 427 VKEAGGVAMILANTEINOEDSIDVHLLPATLIGYTESVLLKAYVNATVKPKARIIF 486
DB 278 -----YANA-----LAVGAT 287
QY 487 VIGSRAPAEVAQFARGPSLANPSILKPDMTAPGVNIIAAWPQLGTLGTPYDSRRV 546
DB 288 DQNNR-----ASFQYGTGL-----NIVAPGVIGQSTVPGN-----RYA 322
QY 547 VMSGTSMSCPHVSGTALIRSAYPNWSPAATKSALMTADLYDRQGAIKDGNKPA 606
DB 323 SLSGTSMATPHVAGVAALVKONPSWSNTQIRQLTSTA-----TSLGNSNQ----- 369
QY 607 IGAGHVNPKA 617
DB 370 FSGGLVNAEA 380
```

RESULT 10
US-10-146-905A-8
; Sequence 8, Application US/10146905A
; Publication No. US20030215906A1
; GENERAL INFORMATION:
; APPLICANT: Boon Leong Lim
; TITLE OF INVENTION: Recombinant Bacillus Proteases and Uses Thereof
; FILE REFERENCE: 9661-035-999
; CURRENT APPLICATION NUMBER: US/10/146,905A
; PRIOR FILING DATE: 2002-03-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-10-146-905A-8

Query Match 6.8%; Score 271.5; DB 12; Length 382;
Best Local Similarity 20.5%; Pred. No. 8.1e-17;
Matches 126; Conservative 63; Mismatches 177; Indels 249; Gaps 17;

```
QY 6 FFLCIIIFLLFCSSSEILQKQTYIVQLHPNSETAKTFASKFDWHLSEFLQEAVLGVVEEB 65
DB 12 FALALIFTMAGSTSS-----AQAAGKSNGEKYIVGFQKTMSTMSAAKKXDVISEK 63
QY 66 EPSRELLSYGSAIEGFAAQLTESEAEILRYSPVAVRPHVLQVQTTYSYKFLGLD 124
DB 64 GKVQKQKY---VDAASATLNKAVKELKDPVAVVEEDHVAH---AYAQSVPYGV 117
QY 125 FGNVWSKRFQGGTIIIGVLDTGVPSPSFDGTGMPSPRKKWKGICQEGESFSSSCN 184
DB 118 IKAPALHSQGYTGNVKNVAVIDSGIDSHPLDKVAGGASM-----ALNNS 157
QY 185 RKLIGARFFIRGHRVANSPEESPNNPREYISARDSTGHGTHASTVGGSSVSMANVL 244
DB 158 -----VPSETNPFQDNNHGHVAGTVA----- 185
QY 245 AQVARGMAPGAHIAVYKVMFNGC-YSSDILAAIDVAIQKVDVLSLGGFFPIPLYDDT 303
DB 186 IGVLF-GVAPASLAVKVLGADGSGQYSWIIINGIEWAIANNMVDVINMSLGG---PS 241
QY 304 IAIGTFRAMERGISVICAAGNNGPIESSVANTAPWSTIGAGTILDRFPFAVVRLANG 363
DB 242 LKAADVKAAGVVVVAAGNEGTSGSS-----STVG----- 273
QY 364 YGESLYPGKGIKNAGREVEVIYVTGGDKGSEFCRLGSLPREIRGKMWICDRGVNGR 423
DB 274 -----YPGK----- 277
QY 424 GEAVKEAGGVAMILANTEINOEDSIDVHLLPATL-IGYTESVLLKAYVNATVKPKARI 482
DB 278 -----YPSVIAVGAVDS----- 289
QY 483 FGGTVIGSRAPAEVAQFARGPSLANPSILKPDMTAPGVNIIAAWPQLGTLGTPYDS 542
DB 290 -----SNQRASFSSVGPGL-----DVNAPGVSIQSTLPGN----- 319
QY 543 VNFTVMSGTSMSCPHVSGTALIRSAYPNWSPAATKSALMTADLYDRQGAIKDGNK 602
DB 320 -KYGAYNGTSMASPHVAGAAALILSKHENWINTQVRSLENTT-----TKLG 365
QY 603 GVFAIGAGHVNPKA 617
DB 366 DSFFYVGKGLINQAA 380
```

RESULT 11
US-10-090-624-31
; Sequence 31, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-090-624-31

Query Match 6.8%; Score 271.5; DB 14; Length 382;

Best Local Similarity 20.5%; Pred. No. 8.1e-17; Indels 249; Gaps 17; Matches 126; Conservative 63; Mismatches 177; Indels 249; Gaps 17;

6 PFLCIIIFLLFCSSSEILQKOTYIVQLHPNSETAKTFASKFDWHLSPLOBAVLGVBESEE 65
12 PALALIFTMARGSTSS-----AQAAGKNGEKKYIVGFKQTMSTMSAAKKDVISEK 63
66 EPSRLLYSYGSAIEGFAAQTESAEILRYSPVAVRDPHVLQVOTTYSYKF-LGLDG 124
64 GGVKQKQFKY---VDAASATLNEKAVKELKDPVAVVEEDHVAH---AVAQSPVYGVQS 117
125 FGNSSGWSKSPFGQGTIGVLDGTGWPSPSPDDTGMPSIPRKWKGCQEGESFSSSSCN 184
118 IKAPALHSQGYTGSNVKAVIDSGIDSHPLKVVAGASM----- 157
185 RKLIGARFFIRGHRVANSPESSPNPREYISARDSTGHGTHASTVGGSSVSMANVLNG 244
158 -----VPSETNPFQDNNSHGTHVACTVA-----ALNNS 185
245 AGVARGMAPGAHIAVYKVCWFNGC-YSSDILAAIDVAIQDKVDVLSLGLGFFPIPLYDDT 303
186 IGV-LGVAPSASLYAVKVLGADGSGQYSWIINGIEWAIANNMVDVINMSLGG---PSGSAA 241
304 IAGTFRAMERGISVICAGNNGPIESSVANTAPWVSTIGAGTLDLDRPPAVVRLANGKLL 363
242 LKAAVDKAVASGVVVAAGNEGTSKSS-----STVG----- 273
364 YGESLYPGKGIKNAGREVEVIVTGGDKSBEFLRGLSPREEIRGKMWICDRGVNCRSEK 423
274 -----YPGK----- 277
424 GEAVKEAGGVAMILANTEINOEDSIDVHLLPATL-IGYTESVLLKAYVNAVTKPKARI 482
278 -----YPSVIANGAVDS----- 289
483 FGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMTAPGVNTIAAMPQNLGPTGLPYDSRR 542
290 -----SNQASPSVSGPEL-----DVMAFGVSIQSTLPGN----- 319
543 VNFTVMSGTSMSCHVSGITILIRSAYPNWSPPAAIKSALMTADLYDRQKAKDGNKPA 602
320 -KYGAYNGTSMASPHVAGAAAILSKHPNWTNTQVRSSLENT-----TKLG 365
603 GVFAIGAGHVNPKA 617
366 DSFYKGLINVOAA 380

RESULT 12
US-10-104-693-2
; Sequence 2, Application US/10104693
; Publication No. US20030118605A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; APPLICANT: Ganshaw, Grant C.
; APPLICANT: Harding, Fiona A.
; APPLICANT: Larenas, Edmund A.
; APPLICANT: Poulos, Ayrookaran J.
; APPLICANT: Sikorski, Elizabeth E.
; APPLICANT: Russell, Elliott P.
; TITLE OF INVENTION: Proteins Producing an Altered Immunogenic
; FILE OF INVENTION: Response and Methods of Making and Using the Same
; FILE REFERENCE: GC683-2
; CURRENT APPLICATION NUMBER: US/10/104,693
; CURRENT FILING DATE: 2002-08-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus amyloiquefaciens
US-10-104-693-2

Query Match 6.8%; Score 271.5; DB 15; Length 382;
Best Local Similarity 20.5%; Pred. No. 8.1e-17; Indels 249; Gaps 17; Matches 126; Conservative 63; Mismatches 177; Indels 249; Gaps 17;

6 PFLCIIIFLLFCSSSEILQKOTYIVQLHPNSETAKTFASKFDWHLSPLOBAVLGVBESEE 65
12 PALALIFTMARGSTSS-----AQAAGKNGEKKYIVGFKQTMSTMSAAKKDVISEK 63
66 EPSRLLYSYGSAIEGFAAQTESAEILRYSPVAVRDPHVLQVOTTYSYKF-LGLDG 124
64 GGVKQKQFKY---VDAASATLNEKAVKELKDPVAVVEEDHVAH---AVAQSPVYGVQS 117
125 FGNSSGWSKSPFGQGTIGVLDGTGWPSPSPDDTGMPSIPRKWKGCQEGESFSSSSCN 184
118 IKAPALHSQGYTGSNVKAVIDSGIDSHPLKVVAGASM----- 157
185 RKLIGARFFIRGHRVANSPESSPNPREYISARDSTGHGTHASTVGGSSVSMANVLNG 244
158 -----VPSETNPFQDNNSHGTHVACTVA-----ALNNS 185
245 AGVARGMAPGAHIAVYKVCWFNGC-YSSDILAAIDVAIQDKVDVLSLGLGFFPIPLYDDT 303
186 IGV-LGVAPSASLYAVKVLGADGSGQYSWIINGIEWAIANNMVDVINMSLGG---PSGSAA 241
304 IAGTFRAMERGISVICAGNNGPIESSVANTAPWVSTIGAGTLDLDRPPAVVRLANGKLL 363
242 LKAAVDKAVASGVVVAAGNEGTSKSS-----STVG----- 273
364 YGESLYPGKGIKNAGREVEVIVTGGDKSBEFLRGLSPREEIRGKMWICDRGVNCRSEK 423
274 -----YPGK----- 277
424 GEAVKEAGGVAMILANTEINOEDSIDVHLLPATL-IGYTESVLLKAYVNAVTKPKARI 482
278 -----YPSVIANGAVDS----- 289
483 FGGTVIGRSRAPEVAQFSARGPSLANPSILKPDMTAPGVNTIAAMPQNLGPTGLPYDSRR 542
290 -----SNQASPSVSGPEL-----DVMAFGVSIQSTLPGN----- 319
543 VNFTVMSGTSMSCHVSGITILIRSAYPNWSPPAAIKSALMTADLYDRQKAKDGNKPA 602
320 -KYGAYNGTSMASPHVAGAAAILSKHPNWTNTQVRSSLENT-----TKLG 365
603 GVFAIGAGHVNPKA 617
366 DSFYKGLINVOAA 380

RESULT 13
US-09-870-122-3
; Sequence 3, Application US/09870122
; Patent No. US20020142009A1
; GENERAL INFORMATION:
; APPLICANT: Regents of the University of Minnesota et al.
; TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
; FILE REFERENCE: 600.450W01
; CURRENT APPLICATION NUMBER: US/09/870,122
; CURRENT FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: US 09/206,898
; PRIOR FILING DATE: 1998-12-07
; PRIOR APPLICATION NUMBER: US 08/589,756
; PRIOR FILING DATE: 1996-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1150
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-09-870-122-3

Query Match 6.7%; Score 269.5; DB 10; Length 1150;
Best Local Similarity 24.2%; Pred. No. 8.3e-16;


```
Matches 128; Conservative 92; Mismatches 201; Indels 107; Gaps 25;
QY 133 KSPFGQGTICVLDGTGWPEPSFDDTGMPS-STPRKWKICQEGESFSSSSNCRKLIGAR 191
Db 117 KAGKAGTVAVIDAG-----FDKNHEAWRLTDKTARYOSKEDLEKA---KKEHGIT 166
QY 192 FFIRGHRVANSPEESPNMPREY:SARDSTGHGHTHTASTVGGSSVSMANVLGNGAGVARGM 251
Db 167 Y---GEWNDKVAYHYDYSKDGTAVDQOE-HGTVSGILSGNAPSTKEPYR-----LEGA 218
QY 252 APGAHIAVYKVCWFGNC--YSSDILAAIDVAIQDKVDVLSLGGFPPI---LYDDTIAI 306
Db 219 MPEAQILLMRVEIVNGLDVARNYAQAIRDAINLGAKVINMSEGNALAYANLPDETCKA 278
QY 307 GTFRAMEGHSVCAAGNG-----PI-----ESSVANT---APWVSTIGAGTLDREF 351
Db 279 FDY-AKSGVSIIVTSAGNDSFSGKTRPLADHPDYGVWGTTPAADSTLTIVASYPDKQL 337
QY 352 PAVURLANG-----KLLYGESLYPGKGIKNAGREVEVIYVTGGKSGEFCLRGSLPRE 404
Db 338 TETVRVTADQCKEMPEVLSTNRPENKAYDIA-----YANRGTKEDDF-----K 382
QY 405 EIRGKMWICDRGVNGRSEKGEAVKEAGGVAMILANTEINOEE-----DSIDVHLLPATLIG 460
Db 383 DVKGKIALIERGDDIDFKIAKAKAGAGVGLIYD---NQDKGFPTELPNVQMPAAFTS 439
QY 461 YTESVLKAYVNATVPKARIIFGGT--VIGRSRAPEVAOFSARGPSLANPSILKXPDMA 518
Db 440 RKDGLLLKD-----NPQKITFNATPKVLPTASGTKLSRFSSWG--LTADGNIKEDIAA 491
QY 519 PGVNIIAAMPQNLGPTGLPYDSRRVFTVMSGTSMSCHPVSGITALIR-----SAYPNWSP 574
Db 492 PGQDILSSVANN-----KYAKLSGTSMSAPLVAGIWMGLLQKQYETQYDPDMP 538
QY 575 A----AIKSAALMTTAD-LYDRQKKAIKDGNKPAGVFAIGAGHVNPOKA 617
Db 539 SERLDLAKVLMSSATLALYDEDEKAYSPRQO-----GAGAVDAKKA 580

RESULT 14
US-10-146-905A-11
; Sequence 11, Application US/10146905A
; Publication No. US20030215906A1
; GENERAL INFORMATION:
; APPLICANT: Boon Leong Lim
; TITLE OF INVENTION: Recombinant Bacillus Proteases and Uses Thereof
; FILE REFERENCE: 9661-035-999
; CURRENT APPLICATION NUMBER: US/10/146,905A
; CURRENT FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion protein between the signal peptide from B. subtilis
; and BPN' protease from B. amyloliquefaciens
US-10-146-905A-11

Query Match 6.6%; Score 266.5; DB 12; Length 382;
Best Local Similarity 20.5%; Pred. No. 2.4e-16;
Matches 127; Conservative 65; Mismatches 168; Indels 261; Gaps 20;
QY 6 FFLCIIF-LFCCSSSSEILQ-----KQYIVQLHPNSETAKTPASKFDWHLSPFQEAVLG 59
Db 12 FALTILFTAFNSNQAAGKNGEKYIVGFQKTMSTMSA-AKKKD-----V 59
QY 60 VEEEEEPSRLLYSYGSAIEGFAAQLTESEAEILRYSPVEVAVRDPHVLQVQVTTYSYKF 119
Db 60 ISEKGGKVKQFKY-----VDAASATLNEKAVKELKXDPVAYVEEDHVAH---AYAQSV 111
QY 120 -LGLDGFNGSWMKSRFGQGTIIIGVLTGTWVPSPSPDDTGMPSIPRKWKICQEGESF 178
```

```
Db 112 PYGYSQIKAPALHSQGYTGSNVKVAVIDSGIDSSHDPDLKVAGGASM----- 157
QY 179 SSSSCNKRKLIGARFFIRGHRVANSPEESPNMPREY:SARDSTGHGHTHTASTVGGSSVMA 238
Db 158 -----VPSETNPFQDNNSHGTHVACTVA----- 180
QY 239 NVLNGAGVARGVAPGAHIAVYKVCWFGNC--YSSDILAAIDVAIQDKVDVLSLGGFPPI 297
Db 181 -ALANSIGVL-GVAPSASLYAVKVLGADGSGQYSWIINGIEWAIANNMDVINMSLGG--- 235
QY 298 PLYDDTTAIGTFRAMEGHSVCAAGNGPIESVANTAPWVSTIGAGTLDREFPAVURL 357
Db 236 PSGSAALKAADVAKAVASGVVVVAAAGNEGTSGSS-----STVG----- 273
QY 358 ANGKLLYGESLYPGKGIKNAGREVEVIYVTGGKSGEFCLRGSLPREIRGKMWICDRGV 417
Db 274 -----YPGK----- 277
QY 418 NGRSEKGEAVKEAGGVAMILANTEINOEEPSIDVHLLPATL-IGYTESVLLKAYVNATVK 476
Db 278 -----YPSVIAVGAVD----- 289
QY 477 PKARIIFGGTVIGRSRAPEVAOFSARGPSLANPSILKXPDMAIPGVNIIAAMPQNLGPTGL 536
Db 290 -----SNQRASFSSVGPGL-----DYMAPGVSIQSTLPGN----- 319
QY 537 PYDSRRVFTVMSGTSMSCHPVSGITALIRSAVFNWSPAAIKSAALMTTADLYDRQKKAIK 596
Db 320 -----KYGAYNGTSMASPHVAGAAALILSKHPNWTNTQVRSLNT----- 361
QY 597 DGNKPAGVFAIGAGHVNPOKA 617
Db 362 --TKLGDSEFFYKGLIINVQAA 380

RESULT 15
US-09-813-408-20
; Sequence 20, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Matrix, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-09-813-408-20

Query Match 6.6%; Score 264.5; DB 11; Length 382;
Best Local Similarity 20.2%; Pred. No. 3.8e-16;
Matches 126; Conservative 68; Mismatches 168; Indels 263; Gaps 18;
QY 6 FFLCIIF-LFCCSSSSEILQKQYIVQLHPNSETAKTPASKFDWHLSPFQEAVLGVVEEEE 65
Db 6 FALALIFTAFNGSTSS-----AQAGKNGEKYIVGFQKTMSTMSAAKKXDVISEK 57
QY 66 EPSEBLLYSYGSAIEGFAAQLTESEAEILRYSPVEVAVRDPHVLQVQVTTYSYKFLGLDGF 125
Db 58 GKGVKQVKFY---VDAASATLNEKAVKELKXDPVAYVEEDHV-----AHAYAQSVPY 107
QY 126 GNSGVMKSRFQOQ-----TIIGVLTGTWVPSPSPDDTGMPSIPRKWKICQEGESF 178
Db 108 GVSQIKAPALHSQGYTGSNVKVAVIDSGIDSSHDPDLKVAGGASM----- 153
QY 179 SSSSCNKRKLIGARFFIRGHRVANSPEESPNMPREY:SARDSTGHGHTHTASTVGGSSVMA 238
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 3, 2004, 00:24:01 ; Search time 376 Seconds
(without alignments)
5564.001 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MEKPFLLCIPLLFCSSES.....NSHNLQVRSPISVTLKTN 775

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09806767/runat_02022004_085957.17887/app_query.fasta_1.967
-DB=N Geneseq 19Jun03 -OFMT=fastap -SUFFIX=xng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09806767 @CNG 1.1 @runat_02022004_085957.17887 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03:*

1: /SID1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SID1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SID1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SID1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SID1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SID1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SID1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SID1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SID1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query #	Score	Match	Length	DB ID	Description
1	4018	100.0	2328	21	AAZ93808	SDD1 sequence enco
2	4009	99.8	2328	21	AAZ93809	SDD1 sequence enco
3	2786.5	69.4	3140	21	AAZ93814	Subtilase (St P2SC
4	2727.5	67.9	2492	21	AAZ93813	Subtilase (Plgsep)
5	2531.5	63.0	3865	21	AAZ93812	Subtilase (Plgsep)
6	1742	43.4	2265	21	AAZ93817	Arabidopsis thalia
7	1742	43.4	2265	24	ABZ13260	Arabidopsis thalia
8	1700	42.3	2637	21	AAZ46158	Arabidopsis thalia
9	1684	41.9	2638	21	AAZ42012	Arabidopsis thalia
10	1642.5	40.9	2434	21	AAZ47609	Arabidopsis thalia
11	1589	39.5	2295	24	ABZ13381	Arabidopsis thalia
12	1589	39.5	2489	21	AAZ42230	Arabidopsis thalia
13	1214	30.2	2402	21	AAZ42039	Arabidopsis thalia
14	1172.5	29.2	2552	15	AAQ73756	Gene encoding melo
15	1095.5	27.3	2448	20	AAZ29905	cDNA encoding a SC
16	1074	26.7	2958	24	ABZ12533	Arabidopsis thalia
17	1070	26.6	2310	25	ABZ42107	Arabidopsis thalia
18	1056.5	26.3	1653	21	AAZ32938	Arabidopsis thalia
19	909	22.6	513445	22	AAI61373	Soybean 318013 reg
20	901	22.4	2451	24	ABZ12259	Arabidopsis thalia
21	806	20.1	7235	20	AAZ29910	Genomic DNA encodi
22	623.5	15.5	740	24	ABQ55482	Arabidopsis thalia
23	448.5	11.2	975	21	AAZ49798	Arabidopsis thalia
24	430	10.7	2532	13	AAQ29134	Encodes RP-III res
25	388	9.7	436	25	ABX61910	Arabidopsis thalia
26	387	9.6	1674	24	ABK78984	Bacillus clausii G
27	345	8.6	1977	18	AAZ85669	Protease coding se
28	332	8.3	1962	18	AAZ85695	Pyrococcus furiosu
29	332	8.3	1962	20	AAZ05929	Hyperthermostable
30	330	8.2	1977	18	AAZ85667	Thermococcus prote
31	330	8.2	1977	20	AAZ05926	WO9856926 Seq ID 1
32	324.5	8.1	1236	20	AAZ05920	Hyperthermostable
33	324.5	8.1	1566	18	AAZ85669	Pyrococcus furiosu
34	322.5	8.0	7156	19	AAV15586	Lactobacillus bulg
35	297	7.4	2539	18	AAZ61454	Streptomyces virid
36	297	7.4	2809	18	AAZ61455	Dhpa-mel chimeric
37	296.5	7.4	4740	25	ABZ58957	Group B Streptococ
38	291.5	7.3	2121	23	ABZ54900	T. yonsei subili
39	291.5	7.3	4941	24	ABN59192	Streptococcus poly
40	289	7.2	528	25	ABX57523	Arabidopsis thalia
41	288.5	7.2	2655	24	ABN67963	Streptococcus poly
42	288.5	7.2	2655	24	ABN70427	Streptococcus poly
43	288.5	7.2	2155561	24	ABN71527	Streptococcus poly
44	287	7.1	343	25	ABX21785	Human GDP-mannose
45	285	7.1	4650	24	ABN71526	Streptococcus poly

ALIGNMENTS

RESULT 1

AAZ93808
ID: AAZ93808 standard; DNA; 2328 BP.

AC AAZ93808;

XX 16-AUG-2000 (first entry)

XX SDD1 sequence encoding a subtilisin like serine protease.

XX SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;
KW crop protection; feed; foodstuffs; ss.

XX Arabidopsis thaliana.

XX Key

Location/Qualifiers

FH


```

Db 1441 ATAAATTTTGGTGGTACGGTGATTTGGAGAGTACAGACACCGAGGTGGCTCAGTTTCA 1500
Qy 501 AlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGly 520
Db 1501 GCTCGAGGACCGAGTTTACCAATCTTCGATACATAAACCCGGATATGATTCTCGGGA 1560
Qy 521 ValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSer 540
Db 1561 GTCAATATCATCTCGGCTTGGCTCAAAATCTAGGACCAACCGGACTTCTTTATGATTC 1620
Qy 541 ArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGly 560
Db 1621 AGAAGAGTTAACTTCACTGTAATGTTCAGGAATTCATGTCTGTCCACATGTTAGCGGA 1680
Qy 561 IleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAla 580
Db 1681 ATCAGTGTCTTATCCGCTCTGCATACCCGAATGCTCTCCAGCTGCATCAATCAATCCGA 1740
Qy 581 LeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLeuAlaIleLysAspGlyAsnLys 600
Db 1741 TTGATGACACACCGGATTTGTACGATCGTCAAGGGAAGCGATAAAGGATGTAACAAA 1800
Qy 601 ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnPro 620
Db 1801 CCAGCCGGTGTGTTGCGATTGGAGCAGGCGCATGTCAATCCGCAAAAGCGATAAACCCG 1860
Qy 621 GlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPhe 640
Db 1861 GGAATTGGTTTACAAACATTAACACAGTGGATTACATAAATCTTACTCTGGACTCTGGATT 1920
Qy 641 ThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArg 660
Db 1921 ACAAGATCAGATATTTTAGCAATCACTCATAGAACGTGAGCTGCATGATGATATGCGG 1980
Qy 661 LysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
Db 1981 AAAAACCCGGTTTTAGTCTCAATTAACCGTCGATAGCCGTGATTTTCAACGTCGAAG 2040
Qy 681 ThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSer 700
Db 2041 ACTACGAGATGATCACAGCGGTGTCACTAACGTTGGGAGTCTTAACTCGATATCTCA 2100
Qy 701 ValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPhe 720
Db 2101 GTGAATGTCAAGGCTCCAGAGGGGATCAAGTATTGTCAATCTTAAGAGACTTGTGTTC 2160
Qy 721 LysHisValAspGlnThrLeuSerTyrArgValTrpPheValLeuLysLysLysAsnArg 740
Db 2161 AAACACGTGGATCAGACGTGAGCTATAGATATGGTTGTATGAGAGAGAAACACAGA 2220
Qy 741 GlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeu 760
Db 2221 GGAGGGAAGTGGCTAGCTTTGCACAGGCGCACTTGACTTGGGTCAACTCTCATATCTG 2280
Qy 761 MetGlnArgValArgSerProIleSerValThrLeuLysThrAsn 775
Db 2281 ATCAGCGAGTTAGAGTCCAAATCTGTAAACCTTGAAGACTAAC 2325

```

RESULT 2

AAZ93809

ID -AAZ93809 standard; DNA; 2328 BP.

XX AC

XX AAZ93809;

XX DT

XX 16-AUG-2000 (first entry)

XX DE

XX SDD1 sequence encoding a subtilisin like serine protease.

XX XX

KW SDD1; serine protease; subtilisin; transgenic plants; dry weight;

KW stomata; sugar; water; protein; CO₂; H₂O; CO₂; H₂O;

KW crop protection; feed; foodstuffs; ss.

XX XX

```

OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT CDS 1..1476
FT /*tag= a
FT /product= SDD1 subtilisin-like serine protease.
XX
FN WO200022144-A2.
XX
PD 20-APR-2000.
XX
PF 12-OCT-1999; 99WO-EPO7633.
XX
PR 12-OCT-1998; 98EP-0119244.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
PI Berger D, Altmann T;
XX
DR WPI; 2000-317995/27.
XX P-PSDB; AAY83301.
XX
PT Novel recombinant DNA molecules encoding subtilisin-like serine
PT protease, useful for producing transgenic plants with altered stomata,
PT lower water consumption and enhanced diseased resistance
XX
PS Claim 1; Page 66-68; 101pp; English.
XX
CC Sequences encoding SDD1, a subtilisin-like serine protease, can be
CC used to produce transgenic plants with altered stomata
CC characteristics. These plants exhibit improved freshness,
CC increased dry weight, reduced leaf temperatures, reduced water loss
CC and lower water consumption and for enhancing the sugar and/or
CC protein content of plant leaves, modulating CO2 uptake into and H2O
CC release from leaves, for sustained photosynthesis under high
CC intensity conditions or for the improvement of disease resistance
CC of plants. The transgenic plants and cells of such plants are useful
CC in the preparation of feed, food or additives. This sequence is
CC identical to the one given in GENESEQ record AAZ93808 with the
CC exception of a single point mutation C/G to T/A at position 1476
CC which introduces a TGA codon and results in premature termination
CC of the SDD1 polypeptide.
XX
SQ Sequence 2328 BP; 635 A; 487 C; 577 G; 629 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 1.73e-311 Length: 2328
Score: 4009.00 Matches: 774
Percent Similarity: 99.87% Conservative: 0
Best Local Similarity: 99.87% Mismatches: 1
Query Match: 99.78% Indels: 0
DB: 21 Gaps: 0

```

US-09-806-767-2 (1-775) x AAZ93809 (1-2328)

```

Qy 1 MetGluProLysProPhePheLeuCysIleIlePheLeuPheCysSerSerSerSer 20
Db 1 ATGGAACCCAAACCTTTCTCTCGATTATCTTTCTATTCTTCTTCGTC 60
Qy 21 GluIleLeuGlnLysGlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLys 40
Db 61 GAGATCTCTGCAAGACGACATTACATTGTCAGTTTCATCTATAGCAACCGCTAAA 120
Qy 41 ThrPheAlaSerLysPheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyVal 60
Db 121 ACCTTTGCTTCARAGTTTGATTGGCATCTTTCTTTCTCCAGAGCGGTTTAGGTGT 180
Qy 61 GluGluGluGluGluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGlu 80
Db 181 GAAGAAGAGAGGAGGAGCGCTTCTTCTCGACTTCTCTACTCTCTATGGCTCTCGATTGAA 240
Qy 81 GlyPheAlaAlaGlnLeuThrGluSerGluAlaGluLeuLeuArgTyrSerProGluVal 100

```

Db 241 GGATTTGCTGCTCAGTTGACTGAATCAGAGCCGAGATACCTGAGATATTCACCTCAAGTT 300
Qy 101 ValAlaValArgProAspHisValLeuGlnValGlnThrThySerThyIysPheLeu 120
Db 301 GTTGCAAGTGAAGCTGACCATGTTCTTCAAGTTTCAACACCATCTACTCTTCAAGATTTCTTG 360
Qy 121 GlyLeuAspGlyPheGlyAsnSerGlyValTyrSerLysSerArgPheGlyGlnGlyThr 140
Db 361 GGCCTCGACGGTTTGGAACTCCGGGTGATGGTCTAAATCTCGTTTGGTCAAGGCACA 420
Qy 141 IleIleGlyValLeuAspThrGlyValTyrProGluSerProSerPheAspAspThrGly 160
Db 421 ATATCGCGGTGCTTGATCTGAGTTTGGCCCTGAAAGTCTTAGCTTTCACGATACCGGA 480
Qy 161 MetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlySerPheSerSer 180
Db 481 ATCCCTTCGATTCACCGAAATGGAAGGATTTGCCAAGAAGGAGAAATTTTCAGTTCT 540
Qy 181 SerSerCysAsnArgLysIleGlyAlaArgPhePheIleArgGlyHisArgValAla 200
Db 541 TCAGCTGTAAACCGGAAGCTAATCGGTGCTAGATTCTTTCATCAGAGGACACCGTGTGCT 600
Qy 201 AsnSerProGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThr 220
Db 601 AATTCACGAGGAATCACAACATGCTCGTGAATACATTTCCGCAAGAGATTCAACG 660
Qy 221 GlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaVal 240
Db 661 GGACACGGGACTCACACCGCTCAACAGTTGGTGGATCCTCTGTTTCGATGCGAATGTT 720
Qy 241 LeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyr 260
Db 721 CTTGGCAATGAGCTGGTGTGGCTGGTGGATGGCTCTGGAGCTCACATTCGAGTCTAT 780
Qy 261 LysValCysTrpPheAsnGlyCysTyrSerSerAspIleLeuAlaAlaIleAspValAla 280
Db 781 AAGTCTGTTGGTTCATAGTTGTTACAGCTCTGCATCTAGACAGTATAGATGAGCG 840
Qy 281 IleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyr 300
Db 841 ATTCAGATTAAGTCGATGTTCTTTCGCTTTCCTTGGCGTTTCCCTATTCCTTGTAT 900
Qy 301 AspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCys 320
Db 901 GATCACCAATCGCATTTGGAACATTCGAGCCATGGAACCGGTATATCTGTAATCTGT 960
Qy 321 AlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAlaAsnThrAlaProTyrValSer 340
Db 961 GCAGCTGGTAAACACGGTCCATCGAAAGCTCTGTTCGAAACACAGCTCTTGGGCTCA 1020
Qy 341 ThrIleGlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAlaAsnGly 360
Db 1021 ACCATTGGCGAGCAGCGCTTGATCGAAGATTTCCTGCTGGTCAGATTAGCCACGGA 1080
Qy 361 LysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGlu 380
Db 1081 AAGCTTCTCTATGAGAGTCAATCTATCCGGAAAGGTATAAAGATGCGCGGAGAG 1140
Qy 381 ValGluValIleTyrValThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySer 400
Db 1141 GTTGAGGTGATTATCGTCACAGGAGGAGATAAGGAAGTGAAGTCTCTGTTGAGAGGTCA 1200
Qy 401 LeuProArgGluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArg 420
Db 1201 CTTCCAGAGAAGAATCCGAGGCAAAATGGTGTATTTGTGATTCGCGAGTCAATGGAGA 1260
Qy 421 SerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThr 440
Db 1261 TCGGAGAAAGGAGAGCGGTTAAAGAGCTCGAGGAGTTGCAATGATCTTAGCCAATACA 1320
Qy 441 GluIleAsnGlnGluAspSerIleAspValHisIleLeuLeuProAlaThrLeuIleGly 460

Db 1321 GAGATCAACCAAGAGAAGATTCATTGACGTTTCATCTCTTACCAGCTACATTTGATTGTT 1380
Qy 461 TyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArg 480
Db 1381 TACACTGAGTCACTCTCTTCTGAAGGCTTATGTTAATGCCACGCTGAACCAAGGCGCG 1440
Qy 481 IleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSer 500
Db 1441 ATAATTTTGGTGGTACGCTGATTGGAGGTCATGAGCACCGGAGGTGCTCAGTTTTC 1500
Qy 501 AlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGly 520
Db 1501 GCTCGAGGACCGAGTTTAGCCAAATCTTCGATCTAAACCCGATATGATTGCTCCGGGA 1560
Qy 521 ValAsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGlyLeuProTyrAspSer 540
Db 1561 GTCATATCATTTGGGCTTGGCCTCAAAATCTAGGACCAACCGACTTCCTTATGATTCA 1620
Qy 541 ArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGly 560
Db 1621 AGAAGAGTTAACTTCACTGTAATGTCAGGAATCTCAATGTCTTGTCCACATGTTAGCGGA 1680
Qy 561 IleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAla 580
Db 1681 ATCACTCTCTTATCCGGTCTGCATACCCGAACTGGTCTCCAGCTGCAATCAATCCGCA 1740
Qy 581 LeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLys 600
Db 1741 TTGATGACACAGCGGATTTGTACGATCTCAAGGGAAGCGATAAAGATGTTAACA 1800
Qy 601 ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnPro 620
Db 1801 CCACCGGTGTTTGGATTGGAGCAGGCGCATGTGAATCCGCAAAAGGCGATAAACC 1860
Qy 621 GlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPhe 640
Db 1861 GGATTTGGTTTACACATTCACACAGTGGATTTACATAACTTACCTCTGCATCTTGGATT 1920
Qy 641 ThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArg 660
Db 1921 ACAAGATCAGATATTTTAGCAATCACTAAGACCGTGAAGTGAATGGAATATTCGGG 1980
Qy 661 LysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
Db 1981 AAAAACCGGTTTGTAGTCTCAATACCGTCGATGCGGTGATTTTCAAACGTTGCAAG 2040
Qy 681 ThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSer 700
Db 2041 ACTCGGAGATGATCACAAAGCGTGTCACTAACGTTGGGAGTCTCTAACTCGATATCTCA 2100
Qy 701 ValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPhe 720
Db 2101 GTGAATGTCAAGGCTCCAGAGGGATCAAGATTTATTGTCAATCTTAAGAGACTTGTGTT 2160
Qy 721 LysHisValAspGlnThrLeuSerTyrArgValTyrPheValLeuLysLysLysAsnArg 740
Db 2161 AAACACGTGGATCAGACGCTGATAGATGATGTTTGTATTGAAGAGAAACACAGA 2220
Qy 741 GlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeu 760
Db 2221 GGAGGAAGGTGGCTAGCTTTTGCAAGGCGAGTTGACTTGGTCAACTCTCATAATCTG 2280
Qy 761 MetGlnArgValArgSerProLysSerValThrLeuLysThrAsn 775
Db 2281 ATGCAGCGAGTTAGAAGTCCAATCTCTGTAACCTTGAAGACTAAC 2325

RESULT 3

AAZ93814

ID AAZ93814 standard; DNA; 3140 BP.

XX

AAZ93814;

XX

16-AUG-2000 (first entry)

```

XX Subtilase (St_P2Sca) of Solanum tuberosum.
DE
DE SDD1; serine protease; subtilisin; transgenic plants; dry weight;
KW stomata; sugar; water; protein; CO2; H2O; CO2; H2O;
KW crop protection; feed; foodstuffs; ss.
XX
XX Solanum tuberosum.
OS
XX
XX Key Location/Qualifiers
XX CDS 1..2301
XX /tag= a
XX /product= Subtilase St_P2Sca
XX
XX W0200022144-A2.
XX
XX 20-APR-2000.
XX
XX 12-OCT-1999; 99WO-EP07633.
XX
XX 12-OCT-1998; 98EP-0119244.
XX
XX (PIAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.
XX (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
XX
XX Berger D, Altman T;
XX
XX WPI; 2000-317995/27.
XX P-PSDB; AAY83304.
XX
XX Novel recombinant DNA molecules encoding subtilisin-like serine
XX protease, useful for producing transgenic plants with altered
XX protease, useful for producing transgenic plants with altered
XX stomata,
XX
XX Claim 1; Page 85-88; 101pp; English.
XX
XX Sequences encoding SDD1, a subtilisin-like serine protease, can be
XX used to produce transgenic plants with altered stomata
XX characteristics. These plants exhibit improved freshness,
XX increased dry weight, reduced leaf temperatures, reduced water loss
XX and lower water consumption and for enhancing the sugar and/or
XX protein content of plant leaves, modulating CO2 uptake into and H2O
XX release from leaves, for sustained photosynthesis under high
XX intensity conditions or for the improvement of disease resistance
XX of plants. The transgenic plants and cells of such plants are useful
XX in the preparation of feed, food or additives. This subfamily of
XX Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
XX (See AAZ93808).
XX
XX SQ Sequence 3140 BP; 939 A; 627 C; 690 G; 882 T; 2 other;
XX
XX Alignment Scores:
XX Pred. No.: 2,49e-213 Length: 3140
XX Score: 2786.50 Matches: 524
XX Percent Similarity: 82.48% Conservative: 107
XX Best Local Similarity: 68.50% Mismatches: 125
XX Query Match: 69.35% Indels: 9
XX DB: 21 Gaps: 5
XX
XX US-09-806-767-2 (1-775) x AAZ93814 (1-3140)
XX
XX 7 PheLeuCysIleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLysGln 26
XX
XX 16 TTTCTATGCTTTTACTATGTTTGTTCATACAGCTCAAGATTG-----CAA 66
XX
XX 27 ThrTyrlleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPhe 46
XX
XX 67 ACTTACATAGTTCAGTTACATCCATGAGGAGCAAGACCCCTTTTACTCTTAAACTA 126
XX
XX 47 AspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGlu 66
XX
XX 127 CAATGGCACCTTCTTCTTCTTCAAAAGCA-----GTTCTCTGGAGAACAGAC 177

```

```

QY 67 ProSerSerArgLeuLeuTyrrSerTyrrGlySerAlaIleGlnGlyPheAlaGlnLeu 86
DB 178 TCGCTTCTCGTCTTTTGTACTCTTACCATCTCCGAGTGAAGGTTTGGAGCTCGACTC 237
QY 87 ThrGluSerGluAlaGluIleLeuArgTyrrSerProGluValValAlaValArgProAsp 106
DB 238 ACTGAAGATGAGTTGAGTTGTTAAGGAATCTAATGATGTGTGTGTGTGTGTGTGTGAG 297
QY 107 HisValLeuGlnValGlnThrThrTyrrSerTyrrLysPheLeuGlyLeuAspGlyPheGly 126
DB 298 AGGAGGCTTGAATTCAGACTACTTATTTACAAAGTTCTTGGGATTAAAGTCCA---ACG 354
QY 127 AsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAsp 146
DB 355 AGAGAAGGAGCTTGTTCAGAGTCTGGATTGGTTCAGGGGGCGATCAITGGAGTGTGGAT 414
QY 147 ThrGlyValTrpProGluSerProSerPheAspThrGlyMetProSerIleProArg 166
DB 415 ACTGGAGTTTGGCCAGAAAGTCCAAGTTTGTATCATGGGATGCCACCTGCTCCACAG 474
QY 167 LysTrpLysGlyIleCysGlnGluGlySerPheSerSerSerSerSerCysAsnArgLys 186
DB 475 AAGTGGAGGGTGTCTGCCAAGGAGGACAGGATTTTAATCTTCTAGTTGTAAATCGCAAG 534
QY 187 LeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAenSerProGluGluSer 206
DB 535 CTATTTGTGCAAGGTTTTTTCAGAAAGGACATCGTGTGGCT---TCAATGACATCATCA 591
QY 207 ProAsnMetProArgGluTyrrIleSerAlaArgSerThrGlyHisGlyThrHisThr 226
DB 592 CCAGATGAGTGGAGGAAATATGTCTGCCACGGGATTTCCCATGGCCATGGTACACATACA 651
QY 227 AlaSerThrValGlyGlySerSerValSerMetAlaAenValLeuGlyAsnGlyValagly 246
DB 652 GCATCCACTGCTGGAGGAGCTGCAGTTCCATTCCTGCTGCTCGGAAATGGAGCAGGG 711
QY 247 ValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrrLysValCysTrpPheAsn 266
DB 712 GAGGCTCAGGAGTGGCCCGGGTGGCCACATTCGAATATATAAAGTATGCTGTTTCACT 771
QY 267 GlyCysTyrrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAsp 286
DB 772 GGTGTTTACAGCTCTGATATACCTTGCAGCAATGGATGGCCATCAGAGATGGAGTAGAC 831
QY 287 ValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrrAspAspThrIleAlaIle 306
DB 832 ATATTGTCACTCTCATCTGGTGGCTTCCTATTCCACTTATGATGATGATGATGATGAT 891
QY 307 GlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGly 326
DB 892 GGAAGTTTCCGAGCCTGAGCATGGAGATTTTCAGTTATATGTGTGCTGAGGGAATAATGGA 951
QY 327 ProfileGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyValaglyThr 346
DB 952 CCAATCCAAAGTTCAGTAGCCACCGTGTCTCTTGGATTGGCCACTATTGGTGTGTAGCACA 1011
QY 347 LeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrrGlyGlu 366
DB 1012 CTTGACAGGAGATTTCCAGCGTTCAGTTTCAGTTAGGCAACGGAAGTTCTCTGACGGAGAA 1071
QY 367 SerLeuTyrrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrrVal 386
DB 1072 TCCTTGTACCTCGGGAAGAAATTCCTAGCTCTCAGAAAGATCTTGAGATCGTTTATGTA 1131
QY 387 ThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIle 406
DB 1132 AAGGATAGGACACAGGAGGAGTGAATTTTGTCTTGAGAGGATCGCTATCAAAGCACAGTGC 1191
QY 407 ArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAla 426
DB 1192 CCAGGGGAAAATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1251
QY 427 ValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGlu 446

```

```

Db      1252 GTGAAGGAGGAGTGGTCTCCATGATCTTAGCAATACAGCAATAATATGAGGAA 1311
Qy      447 AspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyThrThrGluSerValLeu 466
Db      1312 GATTCCATTGATCTCCATCTCCAGCAAGCTTGATTGGCTTCGATGAATCAATCAAA 1371
Qy      467 LeuLysAlaThrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyThr 486
Db      1372 TTACAAACATACCTGATCAACAAAGAACCAACAGCTCGATTATATTTGGAGAACG 1431
Qy      487 ValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeu 506
Db      1432 GTAATAGGAAGCTAGACACCTCGCAGTAGCTCAGTTTCTCAAGGGGCGCAAGCTAT 1491
Qy      507 AlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAla 526
Db      1492 ACTGATCCITTCATTCTCAACCTGATTGATTGCTCCAGGGGTAACATAATTCGGCT 1551
Qy      527 TrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgArgValAsnPheThr 546
Db      1552 TGGCCCAAAACCTAGGCCCTGCTTCCCGAAGATTTCAGAGAGATAAATTTCAT 1611
Qy      547 ValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArg 566
Db      1612 GTTATGTGAGGACCTCAATGGCATGCTCTTCAGGAGATTTCAGAGAGATAAATTTCAT 1671
Qy      567 SerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAsp 586
Db      1672 TCAGCTCATCTTAATGGACTCCAGCAGCAATAAGATCCGATTAAATGACCACTGCAGAT 1731
Qy      587 LeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAlaGlyValPheAla 606
Db      1732 ACAGCTGATCATATGGGAAACCAATCATGGATGGAGATGCACAGCTAAACTTTTGCA 1791
Qy      607 IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIle 626
Db      1792 GCTGGAGCTGGACAGCGTAACCTCGGAAGAGCCATGCTCGATTGATATATGACATC 1851
Qy      627 GlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeu 646
Db      1852 CAGTTTGATGAATATATCATCTATCTTTGACATATCGGATACAGAAATTCGAGTCTTC 1911
Qy      647 AlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSer 666
Db      1912 AGCATTACTATAGGAATGTCAGTGCATGATCATTTTACAGAACCAACAGGGGTTTCAGC 1971
Qy      667 LeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThr 686
Db      1972 CTAAATTTACCCCTCAATTTCAATAACTTTTCAGAGCAGGAATGACTAGAAGATAATCAAG 2031
Qy      687 ArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaPro 706
Db      2032 AGGAGAGTAAATAATGGGGAACCCCTTAATCTATTTACTCAGTTGACATTTGAGGACCT 2091
Qy      707 GluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGlnThr 726
Db      2092 GAGGAGTCAAAGTGAGAGTGAACCCAGCTGCTGATATTTAAACATGTGACCAACAGC 2151
Qy      727 LeuSerTyrArgValTppPheValLeuLysLysAsnArgGlyLysValAlaSer 746
Db      2152 TTAAGCTATAGAGTTGGTTTATA--TCACGAAGAAGAAATAGAGTCTTAAAGGATCAGC 2208
Qy      747 PheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSer 766
Db      2209 TTTGAGAGGGGCAATTGACATGTTTCAATGTAGGAACAAAGCCAGGAAGTTAAAGT 2268
Qy      767 ProfileSerValThr 771
Db      2269 CCTATTTCGGTCACA 2283

```

RESULT 4
AAZ93813

```

ID      AAZ93813 standard; DNA; 2492 BP.
XX      AC      AAZ93813;
XX      DT      16-AUG-2000 (first entry)
XX      DE      Subtilase (Pigesp) of Solanum tuberosum.
XX      KW      SD1; serine protease; subtilisin; transgenic plants; dry weight;
KW      stomata; sugar; water; protein; CO_2; H_2O; CO2; H2O;
XX      crop protection; feed; foodstuffs; ss.
XX      OS      Solanum tuberosum.
XX      FH      Key      Location/Qualifiers
FT      CDS      3..2492
FT      FT      /*tag= a
FT      FT      /product= Subtilase Pigesp
XX      PN      WO200022144-A2.
XX      PD      20-APR-2000.
XX      PF      12-OCT-1999; 99WO-EF07633.
XX      PR      12-OCT-1998; 98EP-0119244.
XX      PA      (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX      FA      (PLAN-) PLANTEC BIOTECHNOLOGIE GMBH.
XX      PI      Berger D, Altmann T;
XX      DR      WPI; 2000-317995/27.
XX      DR      P-PSDB; AAY83303.
XX      PT      Novel recombinant DNA molecules encoding subtilisin-like serine
PT      protease, useful for producing transgenic plants with altered stomata,
PT      lower water consumption and enhanced diseased resistance
XX      PS      Claim 1; Page 79-82; 101pp; English.
XX      CC      Sequences encoding SDD1, a subtilisin-like serine protease, can be
XX      used to produce transgenic plants with altered stomata
XX      characteristics. These plants exhibit improved freshness,
XX      increased dry weight, reduced leaf temperatures, reduced water loss
XX      and lower water consumption and for enhancing the sugar and/or
XX      protein content of plant leaves, modulating CO2 uptake into and H2O
XX      release from leaves for sustained photosynthesis under high
XX      intensity conditions or for the improvement of disease resistance
XX      of plants. The transgenic plants and cells of such plants are useful
XX      in the preparation of feed, food or additives. This subtilease of
XX      Solanum tuberosum is a homologue of SDD1 of Arabidopsis thaliana
XX      (See AAZ93808).
SQ      Sequence 2492 BP; 766 A; 480 C; 511 G; 734 T; 1 other;

```

Alignment Scores:
Pred. No.: 9,74e-209 Length: 2492
Score: 2727,50 Matches: 521
Percent Similarity: 79.22% Conservative: 108
Best Local Similarity: 65.62% Mismatches: 132
Query Match: 67.88% Indels: 33
DB: 21 Gaps: 5

US-09-806-767-2 (1-775) x AAZ93813 (1-2492)

Qy 6 PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLys 25
Db 102 TATTTCTCTGT-----TTTCTACTCTGTTTATCCCTGCTACAGCTCAAAATTG 155
Qy 26 GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThr---PheAlaSer 44
Db 156 CAACTTATATAGTACAAATTATACCATCCACATGCATCAACAGAGACCCCTTTAGTTCT 215

Db 876 TTTTAAAGCTTTAAATTTCTATGTTCCCTCGGAGCCTTONTCTNACTTACTTTTATAC 935
Qy 129 -----GlyValTtpSerLysSerArgPheGlyGlnG1 139
Db 936 TGTCTTTGATCTTTTTTTTCTAAAGGTACTTGGTTAAAGTCTGGATTTTGGTCGAG 995
Qy 139 yThrIleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspAspTh 159
Db 996 CGCGATCATTTGGAGTCTTGTATCTGGAATTTGGCCAGAAAGTCCCAAGTTTGTGATCA 1055
Qy 159 rGlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSe 179
Db 1056 TGGAAATGTCCTCTATTCCAAAGAAATGGAAAGGTTCTGCCAAGAGGAAAGAACTTCAA 1115
Qy 179 rSerSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgVa 199
Db 1116 TTCTTCAAGTTGCAATCGCAAGCTTATTTGGTGCAGAGTTTTCAGATAGGACACATGAT 1175
Qy 199 lAlaAsnSerProGluGluSerProAsnMetProArgGluTyrlleSerAlaArgSpSe 219
Db 1176 GGCATCAAGACATCAAAATCAATAGATTTTATGGAGGATTATGATATCACTCGAGATTC 1235
Qy 219 rThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaAs 239
Db 1236 TCAAGGCCATGTTACATACATACATCTACTGTCAGGGGAGCTCCCGTTTCCAATGGCGAG 1295
Qy 239 nValLeuGlyAsnGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaVa 259
Db 1296 TGTGCTTGGAAATGGAGCAGAGAGGCTCGAGGGATGGCCCTTGGTCTCATATCGCAT 1355
Qy 259 lTyrlleValCysTrpPheAsnGlyCysTyrlleSerSerAspIleLeuAlaAlaIleAspVa 279
Db 1356 ATCAAAAGTTTGTGCTAGTGGTTGTTATAGTTCTGATATATCTTGCAGCAATGGATGT 1415
Qy 279 lAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLe 299
Db 1416 AGCTATTAGAGATGGAGTAGACATATGTTCTTTTCAATGGTGGTTTCCCTGTTCCACT 1475
Qy 299 uTyrlleAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValI1 319
Db 1476 TTATGAGGACTATTGCTATTGGCAGTTTTCAGAGTATGGAACGTGGAATTTTCAGTTAT 1535
Qy 319 eCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpVa 339
Db 1536 ATGTGCTGCGAGAAATATGCTCAATTCCTAAGTTCAGTAGCAAAATGAGGCTCTTGGAT 1595
Qy 339 lSerThrIleGlyValaglyThrLeuAspArgPheProAlaValValArgLeuAla-- 358
Db 1596 TGGCCTATTGGTGTAGCACACTTGACAGAAATTTCCAGCAATATTCAGCTAGG-TA 1654
Qy 358 ----- 358
Db 1655 TGTACATTTTGTCTTAAATGATATTTCGGGTGTTTCCAGCCTAAATATGTGTCCC 1714
Qy 359 -----AsnGlyLysLeuLeuTyrlleGlySerLeuTyrlleProG1 371
Db 1715 TCATTCATATTTTCCACAGGTAAATGCAAGTATGTTATGGAGATCTCTTGACCCGG 1774
Qy 371 yLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrlleValThrGlyGlyAspLy 391
Db 1775 CAACAAGTTCTCAATTTCTCAGAAAGTCTTGAGATTTGTTATCTCAATCAGCGTGATAA 1834
Qy 391 sGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetVa 411
Db 1835 TGGAGTGAATTTTGTATAGAGGTTCTTGCACAGAGCTAAAGTCCATCGGAAATCGT 1894
Qy 411 lIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaG1 431
Db 1895 TGTATGTGATCGTGGAGTTAATGAAGACAGAGAGAAAGGTCAAGTTGTTAAAGAAATCAG 1954
Qy 431 yGlyValAlaMetIleLeuAlaSerThrGluIleAsnGlnGluAspSerIleAspVa 451

```
QY 723 lAspGlnThrLeuSerTyrArgValTrpPheValLeuLysLysLysAsnArgGlyGly 743
Db 3094 TAATCAAGTTTAAGTTACAGAGTTGGTTTATA---TCAAGAGAGAGATTGGACTCA 3150
QY 743 sValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnAr 763
Db 3151 AAGGAGAAGCTTTGCAGAGGACAAATTGATGGATCAACTCCAGAGATAAATACCAGAA 3210
QY 763 sValArgSerProIleSerVal 770
Db 3211 AGTTAGAGTCTTATTCAGTT 3232

RESULT 6
AAC43317
ID AAC43317 standard; DNA; 2265 BP.
AC AAC43317;
XX
XX DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 38825.
DE
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0127452.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 28-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
```


943 AGAGCTTCTGTTGCCAATGTTGCTCCTCGGTTATGACTGTTGGTCTGCTAGTACTTATGAT 1002
Qy
349 AtcArgPheProAlaValAlaValArgLeuAlaAsnGlyValLeuLeuTyrglyGluSerLeu 368
Db
1003 AGAGATTTTCGGGCTTTTCGGAATCTCGGTAAACCGGAACGACTTACCGGTGTTTCGCTG 1062
Qy
369 TyrProGlyLysGlyLysAsnAlaGlyArgGluValGluValileTyrglyValThrGly 388
Db
1063 TATAGCGGTGAGGAATG-----GGACGAGCCGCTTGAATGGTTTATATAAAGG 1116
Qy
389 GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluLeuArgGly 408
Db
1117 AATAGTAGTTCGAGTAATCTTTCTTACCTGGTTCGTTCTGATTCGAGTATTGTTGTTGGG 1176
Qy
409 LysMetValIleCysAspArgGlyValAsnGlyArgSerClyLysGlyGluAlaValLys 428
Db
1177 AAGATTCTGTTGTAGTAGAGTGTAAATGCTAGAGTTGAGAAAGAGAGCTGTGGTTAGA 1236
Qy
429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSer 448
Db
1237 GATCTGCTGTTGTTAGGATGATTAATGGCAATACTGCTCGAGTCGAGGAGGCTGTGG 1296
Qy
449 IleAspValHisLeuLeuProAlaThrLeuIleGlyTyrrThrGluSerValLeuLeuLys 468
Db
1297 GCGGATAGTCAATTTGCTTCCGCGATCGCTGTAGGGAAGAGACTGGTGATTTACTTAGG 1356
Qy
469 AlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIle 488
Db
1357 GAGTATGTTAAGTCAGATCTTAACCAACCGCTCTTCTGTTTAAAGGAACGGTCTT 1416
Qy
489 GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsn 508
Db
1417 GACGTTAAGCGCTCTCTGTTGGTCTGCTTTTAGCTCGAGAGTCTTAATAGTGTACT 1476
Qy
509 ProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaIleAlaIlePro 528
Db
1477 CCTGAATCTGAACCTGATGTTATGCTGCTCGAGTAAATTTTGGCTGCTGCTGCT 1536
Qy
529 GlnAsnLeuGlyProThrGlyLeuProTyrrAspSerArgValAsnPheThrValMet 548
Db
1537 GACGCTATTGCTCTACTGCTTGTGACAAGGACTCTAGGAGGACTCAGTTTCAACATCATG 1596
Qy
549 SerGlyThrSerMetSerCysProHisValSerClyIleThrAlaLeuIleArgSerAla 568
Db
1597 TCAGGTACGTCATGTCATGCCACATCAGTGGTTTACGGGCTCTTTTGAAGGAGCT 1656
Qy
569 TyrProAsnTyrSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrr 588
Db
1657 CACCTGAGTGGATCCGAGTGCTATCAATCAGCTCTCATGACTACAGCTTACGTTCTT 1716
Qy
589 AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla 606
Db
1717 GACAACACCAACGCTCTCTCCATGATGCTGCAGACACACGCTATCTAACCCATATGCT 1776
Qy
607 IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrrAsnIle 626
Db
1777 CACGGCTCGGCCATGTAGATCCCAAGAGCTCTCTCACCAGGCTCTGTCTACGATC 1836
Qy
627 GlnProValAspTyrIleThrTyrrLeuTyrrCysThrLeuGlyPheThrArgSerAspIleLeu 646
Db
1837 TCACCGAGGATACATCAGGTTTGTGCTCTCTAGACTACAGTCGATCATGTTGTT 1896
Qy
647 AlaIleThrHisLys---AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPhe 665
Db
1897 GCGATTGTGAAGCGACCTAGGTTTAACTGCTCGAAGAGTTC---TCAGATCCTGTT--- 1950
Qy
666 SerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIle 685
Db
1951 CAGCTCAACTACCAAGTTTCTCGGTTTGTGTT---GGGGGTAAGAGAGTTGCGGTAC 2007
Qy
686 ThrArgArgValThrAsnValGlySerProAsnSerIleTyrrSerValAsnValLysAla 705
Db
2008 ACTCGGGAAGTAACAAATGTTGGTGCAGCAAGCTCGGTTTACAAGTACGCGTTATGGA 2067

Qy 706 ProGluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGln 725
Db 2068 GCTCCTAGTGTCCGAATCTCTGTTAAACCATCGAAACTTCGTTTAAAGCGTGGAGAG 2127
Qy 726 ThrLeuSerTyrrArgValTyrPheValLeuLysLysAsnArgGlyGlyLysValAla 745
Db 2128 AAGAAGAGGTACACAGTCACGTTTGTAGCAAGAAAGGAGTGTAGTATGACGAACAAGGCT 2187
Qy 746 SerPheAlaGlnGlyGlnLeuThrTyrValAsnSerHisAsnLeuMetGlnArgValArg 765
Db 2188 GAGTTT-----GTTTCATCCTTGGAGCAATCCGAGCAC-----GAAGTGAGA 2232
Qy 766 SerProIleSer 769
Db 2233 AGTCCCGTTGCA 2244
RESULT 7
ABZ13260
ID ABZ13260 standard; DNA; 2265 BP.
XX AC ABZ13260;
XX DT 21-JAN-2003 (first entry)
XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1065.
XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX OS Arabidopsis thaliana.
XX PN WO200216655-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26685.
XX PR 24-AUG-2000; 2000US-227866P.
XX PR 26-JAN-2001; 2001US-264647P.
XX PR 22-JUN-2001; 2001US-300111P.
XX PA (SCRI) SCRIPPS RES INST.
XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX PI Harper JF, Krops J, Wang X, Zhu T;
XX WPI; 2002-304127/34.
PT Identifying a stress condition to which a plant cell has been exposed
and producing plants with increased tolerance to these abiotic stresses
-
PS Claim 144; SEQ ID NO 1065; 577pp + Sequence Listing; English.
XX The invention relates to identifying a stress condition to which a plant
cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
in the plant cell with an array or probes representative of the plant
cell genome; and
CC (b) detecting a profile of expressed polynucleotides in the plant cell
characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 2265 BP; 519 A; 510 C; 542 G; 694 T; 0 other;

Alignment Scores:

Pred. No.: 8.28e-130 Length: 2265

Score:	1742.00	Matches:	370
Percent Similarity:	63.48%	Conservative:	115
Best Local Similarity:	48.43%	Mismatches:	241
Query Match:	43.35%	Indels:	38
DB:	24	Gaps:	15

US-09-806-767-2 (1-775) x ABZ13260 (1-2265)

QY	11	IlePheLeuLeuPheCysSerSerSerSerGluLeuLeuGlnLysGlnThrTrpIleVal	30
DB	52	CTCTCTCTCTCTCCACACACACGCC-----AAAAAACCTACATCATC	96
QY	31	GlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrpHisLeu	50
DB	97	CGCGTCAAT--CACTCCGATAAACCGGAATCTCTCACTACACCGAGATGGTGCACACA	153
QY	51	SerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSerSerArg	70
DB	154	TCTCAACTCAATTCA-----GAAATCATCT	177
QY	71	LeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGlu	90
DB	178	CTTCTCTACACITACACACACCTCTCCATGGCTTCCTCGGTACCTCGACTCCACCGAA	237
QY	91	AlaGlu---IleLeuArgTyrSerProGluValValAlaValArgProAspHisValLeu	109
DB	238	GCGGATTCCT	297
QY	110	GlnValGlnThrThyTyrSerTyrLysPheLeuGlyLeuAspGly---PheGlyAsnSer	128
DB	298	ACACTTCACACTACCGGTACTCCTGAGTTTCTCGGTCTCAATCCGAAATTCGGTGTTCAC	357
QY	129	GlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGly	148
DB	358	GATCTCGGTCTCTCTCT-----AACGGCGTTATCATCGGAGTTTAGATACTGGC	408
QY	149	ValTrpProGluSerProSerPheAspThrGlyMetProSerIleProArgLysTrp	168
DB	409	GATTGGCCCTGAATCTAGAAAGCTTCGATGATCTGATATCGCTGAGATCTCTCTTAATGG	468
QY	169	LysGlyValCysGlnGluGlySerPheSerSerSerSerCysAsnArgLysLeuIle	188
DB	469	AAAGGAGATGTAATCTGGTTCCGATTCGATTCGAATGCCAAGTTGTGTAAAGAGAGCTTATC	528
QY	189	GlyAlaArgPhePheIleArgGlyHisArgValAlaAlaAsnSerProGluGluSerProAsn	208
DB	529	GGAGCTAGAGAGCTTCTCCAAAGAGATTCAAATGCTCTCTGGTGGTGGTTCGAGTAAG	588
QY	209	MetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSer	228
DB	589	-----CGTGAATCTGTTTCTCTCTCGTATGTTGACGGACATGGAACACATACTTCAACT	642
QY	229	ThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAla	248
DB	643	ACCGCGCGGGATCCGCGGTTAGAAACGCTAGCTTCTCTCGGTATACGCCCGCGGTACGGCC	702
QY	249	ArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys	268
DB	703	AGAGGTATGCCCACTCGTCTCGTGTCTCTACTATTAAGTTGTTTGGAGTACTGGTTGT	762
QY	269	TyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeu	288
DB	763	TTTGGATCTGATATACCTACTCTCTATGGATCGAGCTATACTTGTATGTTGTATGTCCTT	822
QY	289	SerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThr	308
DB	823	TCGTTATCTCTTGGTGGTGGTCTCTCGTATTTATCCGCGATAGCATTTGGCATTTGGAGCG	882
QY	309	PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle	328
DB	883	TTTTCGGCTATGGAGAGAGGTGTTTTGTGCTCTGCTGTTATAGTGGTCTTACT	942
QY	329	GluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAsp	348

Db 943 AGAGCTTCTGTTGCCAATGTTGCTCTCTGGGTTATGACTTGTGGTCTGCTGTAGAT 1002
 Qy ArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeu 368
 Db 1003 AGAGATTTTCGGCTTTTCCGAATCTCGGTAAACGGGAAACGACTTACCGGTGTTCGCTG 1062
 Qy TyrProGlyLysGlyLysAsnAlaGlyArgGluValGluValIleTyrValThrGly 388
 Db 1063 TATAGCGGTGAGGAATG-----GGGACCAAGCCGCTCGAATGGTTTTATAATAAAGG 1116
 Qy GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGly 408
 Db 1117 AATAGTAGTTCGAGTAATCTTTGTTACCTGGTTTCGCTTGAATTCGAGTATTTGTCGTGG 1176
 Qy LysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLys 428
 Db 1177 AAGATTGTTGTTGTAGTAGAGGTGTTAATGCTAGAGTTCAGAAAGAGGCTGTGTTAGA 1236
 Qy GluAlaGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSer 448
 Db 1237 GATCGCTGCTGTTTAGGGATGATAATGGCGAATACTGCTCCGAGTCGAGAGGAGCTTGTG 1296
 Qy IleAspValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLeuLys 468
 Db 1297 GCGGATAGTCATTTGCTTCCCGGATCGCTGTAGGGAAGAGACTGCTGTGATTTACTTTAGG 1356
 Qy AlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIle 488
 Db 1357 GAGTAGTTTAAGTCAGATCTTAACCAACCGCTCTCTCTGTTTAAAGGAACGGTTCCT 1416
 Qy GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsn 508
 Db 1417 GACGTTAAGCGCTCTCTGTTGGTTGCTGCTTTAGCTCGAGAGCTCCTAATCTACTGTACT 1476
 Qy ProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpPro 528
 Db 1477 CCTCAATCTGAAGCGCTGATGTTATTTGGTCTCGAGTTAATATTTTGGCTGGTGTCT 1536
 Qy GluAsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsnPheThrValMet 548
 Db 1537 GAGCTATTGGTCTACTGCTCTGCAAGGACTTAGAGGAGCTCAGTTCAACATCATG 1596
 Qy SerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAla 568
 Db 1597 TCAGGTACGTCATGTGTCATCCACACATCAGTGGTGTAGCGGTCTTTTGAAGACAGCT 1656
 Qy TyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyr 588
 Db 1657 CACCTCGAGTGAAGTCCGAGTGTCTCAAAATCAGCTCTCATGACTCAGGTTACTGTTCT 1716
 Qy AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla 606
 Db 1717 GACAACACCAACGCTCTCCATGATGCTGCGACACACAGCCATCTAACCCATATGCT 1776
 Qy IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIle 626
 Db 1777 CACGGCTCGGCCATGTAGATCCCAAGAGGCTCTCTCACAGGTCTTGCTACGACATC 1836
 Qy GluProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeu 646
 Db 1837 TCAACCGAGGAATACATCAGGTTTTTGTGCTCTAGACTACACAGTCGATCACAATGTT 1896
 Qy AlaIleThrHisLys---AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPhe 665
 Db 1897 GCGATTGTGAAGCGACCTACGGTTAATGCTCGAAGAGTTC---TCAGATCTCTGGT--- 1956
 Qy SerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIle 685
 Db 1951 CAGCTCACTACCAAGTTCTCGGTTTCTGTT---GGGGTAAACAGATTGTCGGGTAC 2007
 Qy ThrArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAla 705

[illegible][illegible]

```

Db 1022 GCTATCGCTATTGGAGCTTTTGGAGCTATTGACAGAGGCATATTCTCTCTCTCCGCC 1081
Qy 323 GlyAsnAsnGlyProIleGluSerValAlaAsnThrAlaProTyrValSerThrIle 342
Db 1082 GGTAAACGGAGGTCCTTAAACGGTGACAAATGTTGCTCCGTTGGATGACACAGTC 1141
Qy 343 GlyAlaGlyThrLeuAspArgPheProAlaValAlaValArgLeuAlaAsnGlyLysLeu 362
Db 1142 GGAGCTGGAACAATCATGAGGATTTTCCGGCCAAATGTGAACCTCGTAAACGGGAAGATG 1201
Qy 363 LeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyValGlu 382
Db 1202 ATTCTCGGTGTAGTGTATACGGTGCACCGGCTCTG---GATCCGGTGAATGACCGC 1258
Qy 383 ValIleTyr-----ValThrGlyGlyAsp---LysGlySerGluPheCysLeuArg 398
Db 1259 CTGTGTTACGGTGTAGTGTACTAGCGCGGAGTGTACTCATCTCTCTGTCTGTGAA 1318
Qy 399 GlySerLeuProArgGluLeuLeuArgGlyLysMetValIleCysAspArgGlyValAsn 418
Db 1319 GGCTCGTTGATCCGAATTAGTGAAGGAAGATCGTTCTTTGTGATAGAGAAATCAAT 1378
Qy 419 GlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAla 438
Db 1379 TCTAGAGCAACCAAGGTGAGATCGTACGGAATAATGGAGGCTTGGGGATGATTAGCG 1438
Qy 439 AsnThrGluLeuAsnGlnGluLeuAspSerIleAspValHisLeuLeuProAlaThrLeu 458
Db 1439 AATGGTGTGTTTACGGTGAAGGTTTGTAGTGTGATTGCCACGTGTACCGCGACATCT 1498
Qy 459 IleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLys----- 476
Db 1499 GTTGGTCTCTCGAGGAGATGAGATTCGTAGTATATCTCTCAATCATCCAAATCTCGT 1558
Qy 477 -----ProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArg 492
Db 1559 TCATCGAAACATCCAAACGGTACGATGTTTCAAAGGTACCGACTCGGATCCGACCT 1618
Qy 493 AlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeu 512
Db 1619 GCACCGGTTTTCATCTTCTCTCTGCTGCTGCTGCTTAAATCCAGAGACCGCGAGATCTT 1678
Qy 513 LysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTyrProGlnAsnLeuGly 532
Db 1679 AAACCGGATGATATCGACCTGCTGTTGAATATTCTAGCTGCTTGGCTCACCGGATGCT 1738
Qy 533 ProThrGlyLeuProTyrAspSerArgValAlaAsnPheThrValMetSerGlyThrSer 552
Db 1739 CCATCTGCTGTACTTCTGATACCGGAGAACGTGTTCAACATTTTATCAGGCACCTTCG 1798
Qy 553 MetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTyr 572
Db 1799 ATGCGGTGCGCGACGCTGTCTGTCTAGTCTGCTTAAAGCGCGCTCATCCGATTGG 1858
Qy 573 SerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGly 592
Db 1859 AGTCCGCGCGGATAAGATCAGCATTTGATAACACCGCTTACACGGTTGATACACGCGT 1918
Qy 593 LysAlaIleLysAsp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAla 609
Db 1919 GACCGGATGATGATGATGATCCATCGCAAT---ACATCTTCGTTATGATGATGTTGCTG 1975
Qy 610 GlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProVal 629
Db 1976 GGTATGTTTCAACCAACCAAGCTATGATCCGGATTAGTCTATCATATAACATCTTAT 2035
Qy 630 AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThr 649
Db 2036 GATTATATCAACTCTTGTGTAAATCTTAACCTACCTAGAACCAACATTTGTCAATAACC 2095
Qy 650 HisLysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyr 669

```

```

Db 2096 CGTCCCAAGCGGACTGTGACGGTCAAGACGAGCGGGACACGTCGGGAATTTAACTAC 2155
Qy 670 ProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThr----- 686
Db 2156 CCGTCGTTCTCTGCTGTTTCAGCAGTATGAGAGAGTAAATATGTCACACATTTTCAAT 2215
Qy 687 ArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaPro 706
Db 2216 AGACCTGTGCAATATGAGCGATTTCGGATTTCGGTTTACGAGATTAGGATTAGCGCGCT 2275
Qy 707 GluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGlnThr 726
Db 2276 AGAGGACTACGTTGACGGTTCGAGCGGAGAGCTATCGTTACAGCGGGTGGGCGAGAA 2335
Qy 727 LeuSerTyrArgValTrpPheValLeuLysLysLysAsnArgGlyGlyLysValAlaSer 746
Db 2336 CTGAGT-----TTTGTGTTAGGGTTTAAAGCACACAGAGTGAAGTTGTCAACCA 2383
Qy 747 PheAla-----GlnGlyGlnLeuThrTyrValAsnSerHisAsnLeuMetGln 762
Db 2384 GGAGCGCAAAATGTGGAGACTGGTGCATATAGTTTGGTCAGATGG-----AAACGG 2434
Qy 763 ArgValArgSerProIleSerValThrLeuLys 773
Db 2435 AATGTGCAAGTCCCTTGGTGGTCACTCTTCAA 2467

RESULT 9
AAC42012
ID AAC42012 standard; DNA; 2638 BP.
XX AC AAC42012;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 33960.
XX KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX FN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.

```

PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-01343768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

[illegible]

```
Db 1978 GGTGATGTTACCCACACAGCATGATCGGAGTACTGCTACGATATAACATCTTAT 2037
Qy 630 AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThr 649
Db 2038 GATTATATCAACTTCCTGTGTGTAATCTAATCTACACTGGAACCAACATTTGACGATAACC 2097
Qy 650 HisLysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyr 669
Db 2098 CGTCCCAAGCGGACTCGACGGTTCGAGCGGAGCGGACACGTCGGAATTTGAACACTAC 2157
Qy 670 ProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThr----- 686
Db 2158 CCGTGTCTCTGTCGTTTCACGAGTATGGAGAGTAATATGTCGACACATTTCAATT 2217
Qy 687 ArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaPro 706
Db 2218 AGGACTGTGACAAATGTAGCGATTCCGATTCCGTTTACGAGATTAAGATAGCCGCCCT 2277
Qy 707 GluGlyIleLysValIleValAsnProLysArgLysValPheLysHisValAspGlnThr 726
Db 2278 AGAGGGACTACAGTACAGTGTGAGCGGAGAGAGTATCGTTTCAGACGGTGGGGCAGAAA 2337
Qy 727 LeuSerTyrArgValTyrPheValLeuLysLysLysAsnArgGlyLysValAlaSer 746
Db 2338 CTGAGT-----TTTGTGTGTAGGTTAAGACCAACACAGAGTGAAGTTGTCAACA 2385
Qy 747 PheAla-----GlnGlyGlnLeuThrTyrValAsnSerHisAsnLeuMetGln 762
Db 2386 GGAGCGCAATGTGGAGACTGTCATATGTTTGGTCAGATGG-----AAACGG 2436
Qy 763 ArgValArgSerProIleSerValThrLeuLys 773
Db 2437 AATGTGACAACTCCCTTGGTGTGCACCTTCAA 2469

RESULT 10
AAC47609
ID AAC47609 standard; DNA; 2434 BP.
XX
AC AAC47609;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54454.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
```

```
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 22-JUL-1999; 99US-0145218.
PR 22-JUL-1999; 99US-0145224.
PR 22-JUL-1999; 99US-0145276.
PR 22-JUL-1999; 99US-0145313.
PR 22-JUL-1999; 99US-0145319.
PR 22-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0147038.
PR 02-AUG-1999; 99US-0147204.
PR 02-AUG-1999; 99US-0147302.
PR 02-AUG-1999; 99US-0147192.
PR 02-AUG-1999; 99US-0147260.
PR 02-AUG-1999; 99US-0147303.
PR 02-AUG-1999; 99US-0147416.
PR 02-AUG-1999; 99US-0147493.
PR 02-AUG-1999; 99US-0147935.
PR 02-AUG-1999; 99US-0148171.
PR 02-AUG-1999; 99US-0148319.
PR 02-AUG-1999; 99US-0148341.
PR 02-AUG-1999; 99US-0148565.
PR 02-AUG-1999; 99US-0148694.
PR 02-AUG-1999; 99US-0149368.
PR 02-AUG-1999; 99US-0149175.
PR 02-AUG-1999; 99US-0149426.
PR 02-AUG-1999; 99US-0149722.
PR 02-AUG-1999; 99US-0149723.
PR 02-AUG-1999; 99US-0149929.
PR 02-AUG-1999; 99US-0149902.
PR 02-AUG-1999; 99US-0149930.
PR 02-AUG-1999; 99US-0150566.
PR 02-AUG-1999; 99US-0150884.
PR 02-AUG-1999; 99US-0151065.
PR 02-AUG-1999; 99US-0151066.
PR 02-AUG-1999; 99US-0151080.
PR 02-AUG-1999; 99US-0151303.
PR 02-AUG-1999; 99US-0151438.
PR 02-AUG-1999; 99US-0151930.
PR 02-AUG-1999; 99US-0152363.
PR 02-AUG-1999; 99US-0153070.
PR 02-AUG-1999; 99US-0153758.
PR 02-AUG-1999; 99US-0154018.
PR 02-AUG-1999; 99US-0154039.
PR 02-AUG-1999; 99US-0154779.
PR 02-AUG-1999; 99US-0155139.
PR 02-AUG-1999; 99US-0155486.
PR 02-AUG-1999; 99US-0156659.
PR 02-AUG-1999; 99US-0156458.
PR 02-AUG-1999; 99US-0156596.
PR 02-AUG-1999; 99US-0157117.
PR 02-AUG-1999; 99US-0157753.
PR 02-AUG-1999; 99US-0157865.
PR 02-AUG-1999; 99US-0158029.
PR 02-AUG-1999; 99US-0158232.
PR 02-AUG-1999; 99US-0158369.
PR 02-AUG-1999; 99US-0159293.
PR 02-AUG-1999; 99US-0159294.
PR 02-AUG-1999; 99US-0159295.
PR 02-AUG-1999; 99US-0159329.
PR 02-AUG-1999; 99US-0159330.
PR 02-AUG-1999; 99US-0159331.
PR 02-AUG-1999; 99US-0159637.
PR 02-AUG-1999; 99US-0159638.
PR 02-AUG-1999; 99US-0159584.
PR 02-AUG-1999; 99US-0160741.
PR 02-AUG-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 21-OCT-1999; 99US-0160981.
PR 21-OCT-1999; 99US-0160989.
PR 21-OCT-1999; 99US-0161404.
PR 21-OCT-1999; 99US-0161405.
PR 21-OCT-1999; 99US-0161406.
PR 21-OCT-1999; 99US-0161359.
PR 21-OCT-1999; 99US-0161360.
PR 21-OCT-1999; 99US-0161361.
PR 21-OCT-1999; 99US-0161920.
PR 21-OCT-1999; 99US-0161992.
PR 21-OCT-1999; 99US-0161993.
PR 21-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 8.62e-122 Length: 2434
Score: 1642.50 Matches: 342
Percent Similarity: 59.95% Conservative: 128
Best Local Similarity: 43.62% Mismatches: 281
Query Match: 40.88% Indels: 33
DB: 21 Gaps: 11

US-09-806-767-2 (1-775) x AAC47609 (1-2434)

QY 4 LysProPhePheLeuCysIle-----IlePheLeuLeuPheCysSerSerSer 19
Db 116 AARCCCTTTCTTTTCATAATCTTATCAATCAATCTTCTTCTTCAAGCAGAACAACT 175
QY 20 SerGlutLeuGlnLysGlnThrTyrIleValGlnLeuHisProAsnSerGluThrAla 39
Db 176 ACTCAATCTCTACCAAGAGACCTTATGTTATCCACATG---GATAATCTGCCATGCCT 232
QY 40 LysThrPheAlaSerLysPheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGly 59
Db 233 TTACCTTACACTAATACCTACCTACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 292
QY 60 ValGluGluGluGluGluGluGluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIle 79
Db 293 AAATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 352
QY 80 GluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluLeuLeuArgTyrSerProGlu 99
Db 353 CACGGTTTAGCAGCTCAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 412
QY 100 ValValAlaValAArgProAspHisValLeuGlnValGlnThrThrTyrSerTyrLysPhe 119
Db 413 GTTGTAGCTGTGATACCTGAGACAGATACGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 472
QY 120 LeuGlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGly 139
Db 473 CTGCGTTTAGAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 532
QY 140 ThrIleLeuGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspAspThr 159
Db 533 GTGCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 592
QY 160 GlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSer 179
Db 593 GGTATGTCCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 652
QY 180 SerSerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgVal 199
Db 653 AAACGTAATCGAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 712
QY 200 AlaAsnSerProGluGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSer 219
Db 713 GCACGGGAGAGATCGATGAA-----GAGCTTGAATATAAGTCCCGAGAGACAGA 763
QY 220 ThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerValSerMetAlaAsn 239
```

```
Db 764 GATGGTCACGGGACACACACTGCGCTACTGTAGCTGGCTCACCTGTTAAAGGAGCTAAT 823
Qy 240 ValLeuGlyAsnGlyAlaGlyValAlaAargGlyMetAlaProGlyAlaHisIleAlaVal 259
Db 824 CTTTTGGATTGCTTATGGACAGCTCGAGGATGGCTCAAAAGGCTAGAGTTGCTGCT 883
Qy 260 TyrIysValCysTrpPheAsnGlyCysTyrSerSerAspIleLeuAlaIleAspVal 279
Db 884 TATAAAGCTCTGTTGGTGGGCGAGGCTGTTTCAGATTCACATTTTGTGGGCTGTGATCAA 943
Qy 280 AlaIleGlnAspIysValAspValLeuSerLeuSerLeuGlyPheProIleProLeu 299
Db 944 GCTGTTCTGATGAGTTCAAGTACTCTATATCATTTAGTGGTGGGCTCTCTACTTAT 1003
Qy 300 TyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluAargGlyIleSerValIle 319
Db 1004 TCTAGAGACAGTTTGTCTATAGCAACATTTGGAGCAATGGAGATGGAGTTTCTGTTTCG 1063
Qy 320 CysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpVal 339
Db 1064 TGTTCTGCGGTAATGGAGTCTGATCCGATTAGTCTCACTAATGTTCTCCATGGATC 1123
Qy 340 SerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValValAargIleAlaAsn 359
Db 1124 ACAACAGTTGTCGCAAGTACTATGATAGATGATTTTCCAGCAACAGTGAAGATAGGAAT 1183
Qy 360 GlyIysLeuLeuTyrGlyGluSerLeuTyrProGlyIysGlyIleIysAsnAlaGlyArg 379
Db 1184 ATGAGAACATTCAAAGAGTGTCACTTTCAAAAGGCGAAGACAGTTTGTGCTAAGATAAAA 1243
Qy 380 GluValGluValIleTyrVal-----ThrGlyGlyAspIysGlySerGluPheCys 396
Db 1244 CAGTATCCTCTGTTTACTTAGGAAGGAATGCAAGTAGTCTGATCCACCTCGTCTGT 1303
Qy 397 LeuArgGlySerLeuProArgGluGluIleAargGlyIysMetValIleCysAspArgGly 416
Db 1304 CTAGATGGAGCTTTGGATCGCGCCATGTAGCGGAAAGATCGTGATATGCGACCGCGT 1363
Qy 417 ValAsnGlyArgSerGluIysGlyGluAlaValIysGluAlaGlyValAlaMetIle 436
Db 1364 GTTACTCCAGTGTGCAAAAGGTCAGGTTGTGAAAAGAGCTGGTGGAAATGGGATGTT 1423
Qy 437 LeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeuLeuProAla 456
Db 1424 TTTAACTAACACTGCAACAATGTGGAAGACTGTTGTCAGATTCCTCATGTCTTCAGCT 1483
Qy 457 ThrLeuIleGlyTyrThrGluSerValLeuLeuIysAlaIleTyrValAsnAlaThrValIys 476
Db 1484 GTTGCAGTTGAGAGAAAGAGGTAACTAATCAACAGTACGCGATGACGAGTAAATAA 1543
Qy 477 ProIysAlaArgIleIlePheGlyThrValIleGlyArgSerArgAlaProGluVal 496
Db 1544 GCGACAGCGAGTTTAGAGATTCTTTGGAACAAGAATTTGTTATCAAACTTCACAGTTGA 1603
Qy 497 AlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuIysProAspMet 516
Db 1604 GCAGCGTTCTCTCAAGAGACCAAAATTTCTCTTTGGAGATCTTGAACACGAGCTTG 1663
Qy 517 IleAlaProGlyValAsnIleIleAlaIleTrpProGlnAsnLeuGlyProThrGlyLeu 536
Db 1664 TTGGCTCCAGAGTGAATATCTTTCGAGCTTGGAGATGCGACATGCGACATGAGTTTA 1723
Qy 537 ProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCysPro 556
Db 1724 TCATCTGATCCAGGAGGGTTAAGTTCAATATATCTCTGGAACCTTCATGTCATGCTCT 1783
Qy 557 HisValSerGlyIleThrAlaLeuIleAargSerAlaTyrProAsnTrpSerProAlaAla 576
Db 1784 CATGTAAGTGTGTAGTGTCTTTCATCAATCAAGGCATCCAGATTCGAGTCTCGCAGCA 1843
Qy 577 IleIysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyIysAlaIleIys 596
```

```
1844 ATCAAACTCAGCTCTCATGACAACTGCTTATGTTTCATGACAACTGTTTAAAGCCTCTTACG 1903
Qy 597 Asp-----GlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro 614
Db 1904 GATCATCAGGAGCAGCTCCTTTCATCGCTTATCATCAGCGTGCAGGACATATAGATCCT 1963
Qy 615 GlnIysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyr 634
Db 1964 TTAAGAGCTCAGATCCTGCTGTTGCTACGACATTCGACCTCAAGAGTATTTTGAATTC 2023
Qy 635 LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThr---HisLysAsnVal 653
Db 2024 CTCTGCACTCAAGATTTAAGTCCATCAGCTTAAGGTATTCACAAACATTCACAAACAGA 2083
Qy 654 SerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAla 673
Db 2084 ACCTGCAACACACTCTTGCCAGAATCCGGA---AACTTGAACCTACCCGCGCATATCA 2140
Qy 674 ValIlePheLysArgGlyLysThrThrGluMetIleThr-----ArgArgValThrAsn 691
Db 2141 GCTTTGTTCCAGAGAACACACATGTTAAAGCTATGACACTTAGAAGACAGTGCACCAAT 2200
Qy 692 ValGlySerProAsnSerIleTyrSerValAsnValIysAlaProGluGlyIleLysVal 711
Db 2201 GTTGGTCTCATATTCAAGCTACAAGTTTCTCTCGCCATTCAGAGCGCATCCGTA 2260
Qy 712 IleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyr----- 729
Db 2261 ACTGTCCAGCCCAAAACACTCACTTCACTTCGAGCACCAGAGCTTTCTTACACGGTT 2320
Qy 730 -----ArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAlaSerPhe 747
Db 2321 ACTTTCAGGACAGGTTCCGGATGCAAGGCTGAGTTTGTGTGT----- 2365
Qy 748 AlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerPro 767
Db 2366 -----CTAGTGTGGAAGAGCACT-----ACACATAAAGTTCTGTAGCCCG 2404
Qy 768 IleSerValThr 771
Db 2405 GTTATCATCA 2416
RESULT 11
ID ABZ13381 standard; DNA; 2295 BP.
XX ABZ13381;
AC ABZ13381;
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1186.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216555-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26685.
XX
PR 24-AUG-2000; 2000US-227866P.
XX
PR 26-JAN-2001; 2001US-264647P.
XX
PR 22-JUN-2001; 2001US-300111P.
XX
PA (SCEI ) SCRIPPS RES INST.
XX
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Krepes J, Wang X, Zhu T;
XX
XX WPI; 2002-304127/34.
XX
```



```
Db 1642 GCGGCGCTTCAATCGCTTCACAGATTGGAGCCCTCGGTGATACCAATGCGCAATG 1701
Qy 582 MetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsn----- 599
Db 1702 ATGACACGACTAACCCTCGTCGTAACCTCAACCGCTCGTGTGATCGATGATCCACCGGG 1761
Qy 600 LysProAlaGlyValPheAlaIleGlyValGlyHisValAsnProGlnLysAlaIleAsn 619
Db 1762 AAATCGGCTACGCTATGATGATTCGCGGTCAATTAATTTGGCGCGGTATGAAT 1821
Qy 620 ProGlyLeuValTyrAsnIleGlnProValAspTyrXleThrTyrLeuCysThrLeuGly 639
Db 1822 CCGGCTCTGTCTACGATATACTAATGATGATACATTAAGTTTCTTCTCCATCGGG 1881
Qy 640 PheThrArgSerAspLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeu 659
Db 1882 TAGCGACCAACAGACATCAAGTGAACAGACACCGGTGATGATCGACAGCAGG 1941
Qy 660 ArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePhe----- 676
Db 1942 AAACGCTCTCCGGG---AATTGAAATTATCTTCGATCACGGCGGTGTTTCTACTAAT 1998
Qy 677 LysArgGlyLysThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsn 696
Db 1999 AGAAGAGGATTGTCAGTAAACTGTTATAAGGACGGCGCAATGTCGGCGCAGCTGAG 2058
Qy 697 SerIleTyrSerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLys 716
Db 2059 GCGGTTATCGGGCAGGATAGATCGCCGAGAGAGTGTCAGGTGACAGTGAACACCAT 2118
Qy 717 ArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTyrPheValLeuLys 736
Db 2119 AGGCTTGTTTACTTCGGCCGTTAAGACAGCGAGCTATCGGTTACAGTCAGGTAAAT 2178
Qy 737 LysLysAsn-----ArgGlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTyr 754
Db 2179 ACAAGGAATGTGTGTGGAGAAACAGGTGCTGTGTT-----GGGTGAGTCACGTGG 2232
Qy 755 ValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerValThr 771
Db 2233 TTTGATGTGGAAACACGTG-----GTTCCGAGCCCCATCGGTGGTACCC 2277

RESULT 12
ID AAC42230 standard; DNA; 2489 BP.
XX
AC AAC42230;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 34769.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
```


Qy	187	LeuIleGlyAlaAArgPheIleAArgGlyHisArgValIleAAsnSerProGluGluSer	206
Db	708	ATTATCGAGCAAGATTCTTCGCTAAGGACAACAAGCCGCT-----GTAATCGGAGGA	761
Qy	207	ProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThr	226
Db	762	ATCAACAAAACCGTTGAGTTTCTATCTCTCGTACGCCGATGGACACGGTACTCACACT	821
Qy	227	AlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGly	246
Db	822	TCCTCAACCGCGCTGGCCGTCAGCTTTAAAGCGAGTATGTCGGGTAGCCGCTCCGGT	881
Qy	247	ValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrIysValCysTrp---Phe	265
Db	882	GTAGCCAAAGGTGTGTGCTCCAAAAGCTCGTATCGCCGCTACAAGACTGTGTGAAAGAT	941
Qy	266	AsnGlyCysTyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysVal	285
Db	942	TCCGGTTGTCGATTCGATATCTCGCCGCTTGTATGCCGCTGTATGAGACGGGTGC	1001
Qy	286	AspValLeuSerLeuSerLeuGly-----GlyPheProIleProLeuTyrAspAsp	302
Db	1002	GACGTTATATCGATCTCAATCGGTGGTGAGACGGGATTACTTCGCCGCTATTACTCGAT	1061
Qy	303	ThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAla	322
Db	1062	CCAATCGCTATAGCTGCTCGGCCGCGCTCGAAAGGAATCTTCGCTCTCTTCCTCTGCC	1121
Qy	323	GlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTyrValSerThrIle	342
Db	1122	GGAAACGAAGACCTTAACGGTATGTCACTAGTACGAACCTCGCGCGCTGGGTAAACCACG	1181
Qy	343	GlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAlaAsnGlyLysLeu	362
Db	1182	GGTGCTAGTACAAATCGATCGAAATTTCCAGCGGATGCTATTCTCGCGACGGACATCGT	1241
Qy	363	LeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGlu	382
Db	1242	CTCAGAGGAGTGTCTCTTACGCTGGAGTACCTTTAAAC-----GGTCGTATGTTTCCG	1295
Qy	383	ValIleTyrVal---ThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeu	401
Db	1296	GTGGTTTATCCCGTAAATCGGAATGTCAACAGCTCTCTATGATGGAGAACACCGCTT	1355
Qy	402	ProArgGluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSer	421
Db	1356	GATCCGAAGCAAGTACGGGTAAATAGTATATCTGCATAGAGAGACGAGTCCACCGGTA	1415
Qy	422	GluLysGlyGluAlaValLysGluAlaGlyValAlaMetIleLeuAlaAsnThrGlu	441
Db	1416	GCCAAAGGATTGGTTGTGAAGAAAGCAGGTGGTCTCGGAATGATTCTCGCTAATGGAGCA	1475
Qy	442	IleAsnGlnGluAlaAspSerIleAspValHisIleLeuLeuProAlaThrLeuIleGlyTyr	461
Db	1476	TCTAACCGTGAAGGATTAGTCGGAGATGCTCATCTTATTCAGCGCTGCGGTGGATCA	1535
Qy	462	ThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIle	481
Db	1536	AACGAAGAGATAGAACTCAAGCAATATGCTTCTTCATCCGGAATCCAATGCTTCAATT	1595
Qy	482	IlePheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAla	501
Db	1596	GATTTACAGAGGAACATATAGTTGGGATTAACCGGCTCCGGTTATTGCTTCTTTCTCCGGT	1655
Qy	502	ArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyVal	521
Db	1656	AGAGCAACAAACGGTTTAAAGCCCGGAGATCTCTTAACCGGATTTGATGTGTCCCGAGTT	1715
Qy	522	AsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArg	541
Db	1716	AACATCTCGCCCGCATGACAGACGCTGTGTGGACTCAGAGTTTGGCGTCAGATCCCAAGG	1775
Qy	542	ArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIle	561

Db	1776	AAAACCGGAATTCACATCTCTCTCCGGTACTTCATAGCATGTCTCCTCAGTTAGTGTGGC	1831
Qy	562	ThrAlaLeuIleAArgSerAlaIleValProAsnTyrSerProAlaAlaIleLysSerAlaLeu	581
Db	1836	CGCGGGTCTTCAATCCGCTCATCCAGATGGAGCCCTCCGTGATACGATCGCAATG	1895
Qy	582	MetThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsn	599
Db	1896	ATGACAACGACTAACCTCGTCGATACTAAACCGCTCGTGTGATCGATGAATCCACGGG	1955
Qy	600	LysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn	619
Db	1956	AAATCGCGCTACGCCCTTATGATTACGGGTCCGGTCATTAAATTTGGCCCGGCTATGAAT	2015
Qy	620	ProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGly	639
Db	2016	CCGGGTCTTGTCTACGATATACTAATGATGATTACATTACGTTCTTGTCTCCATCGGG	2075
Qy	640	PheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeu	659
Db	2076	TACGGACCAAGACGATCCAAGTGATAACAAGAACACCGGTGAGATGTCGACGACCAGG	2135
Qy	560	ArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePhe	676
Db	2136	AAACCGCTCCCGGG--AAATTTGAATTATCCTTCGATCACCGCGGTGTTCTCTACTAAT	2192
Qy	677	LysArgGlyLysThrThrGluMetIleThrArgValThrAsnValGlySerProAsn	696
Db	2193	AGAAGAGATTGGTGAGTAAACTGTATTAAGACCGCGACGAAATTCGGGAGGCTGAG	2252
Qy	697	SerIleTyrSerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLys	716
Db	2253	CGCGTTTATCGGGCAGGATAGATGCCGAGAGGATGACGCTGACAGTGAACACCACT	2312
Qy	717	ArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValThrPheValLeuLys	736
Db	2313	AGGCTTGTGTTTACTTCGGCCGTTAAGAGACGAGGCTATGCGGTTACAGTCGCGTTAAT	2372
Qy	737	LysLysAsn----ArgGlyGlyLysValAlaSerPheAlaGlnGlyLysLeuThrTyr	754
Db	2373	ACAAGGAATGTGTGTGGGAGAACACGCTGCTGTGTT-----GGTCAGTCACGTGG	2426
Qy	755	ValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerValThr	771
Db	2427	TTTCTGTTGGGAAACACGTTG-----GTTTCGGAGCCCATCTGTGTGACC	2471
RESULT	13		
AAAC42099			
ID	AAAC42099	standard; DNA; 2402 BP.	
XX	AC	AAAC42099;	
XX	DT	17-OCT-2000 (first entry)	
XX	DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 34284.	
XX	KW	Hybridisation assay; genetic mapping; gene expression control;	
XX	KW	protein identification; signal transduction pathway;	
XX	XX	metabolic pathway; promoter; termination sequence; ss.	
OS	Arabidopsis thaliana.		
XX	PN	EP1033405-A2.	
XX	PD	06-SEP-2000.	
XX	PF	25-FEB-2000; 2000EP-0301439.	
XX	PR	25-FEB-1999; 99US-0121825.	
XX	PR	05-MAR-1999; 99US-0123180.	
XX	PR	23-MAR-1999; 99US-0123548.	
XX	PR	23-MAR-1999; 99US-0125788.	

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135553.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149446.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

D	b	1506	GCTGTTGCATATTTTTTCTTCACAGGACCCTTCAAGCCTCAAGAAGCATTTCTCAAACCT	1565
Q	y	515	AsnMetIleAlaProGlyValAsnIleAlaalaTrpProGlnAsn-----Leu	531
D	b	1566	GACATTACCGCACAGAGAGTCTCATACTTGCTGCATGGACTGGAAACGACTCAAGCAT	1625
Q	y	532	GlyProThrGlyLeuPofTyArgSerArgValAsnPheThrValMetSerGlyThr	551
D	b	1626	TCACTGGAAGGCAAGCCGGCTTCT-----CAGTATAACGTCAATCAGGAAC	1673
Q	y	552	SerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsn	571
D	b	1674	TCCATGGCAGCTCCTCATGTITTCAGCTGTTGCATCTCTGATCAATCACAGCATCCACA	1733
Q	y	572	TrpSerProAlaIleLysSerAlaLeuMetThrThrAla---AspLeuTyrAspArg	590
D	b	1734	TGGGGTCCATCCGGGATCAGATCAGCAATATGACACAGCGACTCAAACAAACACGAC	1793
Q	y	591	GlnGlyLysAlaIleLysAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGly	610
D	b	1794	AAAGTCTTATAACAACAGAAACTGGTGCACACAGCCACACCTTATGACTCTGGAGCAGGA	1853
Q	y	611	HisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAsp	630
D	b	1854	GAACTAAGCTCAACAGCATCAATGCAACAGGACTAGTTCACGAGACTACTGTAACACTGAC	1913
Q	y	631	TyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHis	650
D	b	1914	TACCTGAACCTTCTCTGTATCTATGGATATTAACGTAAACCAATAAAGGCTATGTCAAAA	1973
Q	y	651	-----LysAsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeu	667
D	b	1974	GCTTTCCAGAGAATTTTACTTGGCCTGCAGATTCCAACCTTAGACTTGATCTCCACCATC	2033
Q	y	668	AsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArg	687
D	b	2034	AATTACCCGTCATTTGGAACTCTCTGGATTCAAGAGAAATGGTAGCAAGACAGATTACAAGA	2093
Q	y	688	ArgValThrAsnValGlySerPro---AsnSerIleTyrSerValAsnValLysAlaPro	706
D	b	2094	ACAGTGACCAATGTTGGTGAGAGCGGTGAGCGGTTTACACAGTCAGCGTGGGAACACCA	2153
Q	y	707	GluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGlnThr	726
D	b	2154	CCAGGTTTAACTTCAAGTCAAGCCAGAAAACTCCAATTTACAAAAGATGGTGAGAG	2213
Q	y	727	LeuSerTyrArgValTrpPheValLeuLysLysLysAsnArgGlyLysValAlaSer	746
D	b	2214	TTGACATACCAGGTGATAGTG-----TCTGCTACTGCTTCA	2249
Q	y	747	PheAlaGln-----GlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArg	763
D	b	2250	CTTAAGCAAGATGATTTGGGGCTCTTACTTGGTCTAATGCC-----AAGTACAAG	2300
Q	y	764	ValArgSerProIleSerValThrLeuLysThrAsn	775
D	b	2301	GTCAAGAGCCCAATTGTAATTAGTAGCAGAGATGAC	2336
RESULT 14				
ID		AAQ73756 standard; cDNA to mRNA; 2552 BP.		
XX		AAQ73756;		
AC				
XX		(first entry)		
DT		21-JUL-1995		
XX		Gene encoding melon protease from Cucumis melo L.		
DE		Melon; protease; Cucumis; immunoscreen; antibody; pGEMEX-1; E.coli; ss.		
KW		Cucumis melo (L.) cv Earls Favourite.		
OS				
XX				

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 3, 2004, 02:06:41 ; Search time 101 Seconds
(without alignments)

3386.851 Million cell updates/sec

Title: US-09-806-767-2

Perfect score: 4018

Sequence: 1 MBPKPFLCIIIFLLFCSSSS.....NSHNLQVRSPISVTLKTN 775

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09806767/runat_02022004_085959_17923/app_query.fasta_1.967
-DB=Issued Patents NA -CPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=100 -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09806767@cgn2_1_44@runat_02022004_085959_17929 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	430	10.7	2532	1	US-07-671-376C-4
2	345	8.6	1977	3	US-08-894-818B-6
3	332	8.3	1962	3	US-08-894-818B-34
4	332	8.3	1962	3	US-09-445-472-15
5	330	8.2	1977	3	US-08-894-818B-2
6	330	8.2	1977	4	US-09-445-472-11
7	324.5	8.1	1236	4	US-09-445-472-2
8	324.5	8.1	1566	3	US-08-894-818B-4
9	297	7.4	2539	3	US-09-000-016-3
10	297	7.4	2539	4	US-09-514-340-3
11	297	7.4	2809	3	US-09-000-016-1
12	297	7.4	2809	4	US-09-514-340-1

13	271.5	6.8	1494	3	US-09-255-502-1	Sequence 1, Appli
14	271.5	6.8	1497	6	5472855-1	Patent No. 5472855
15	271.5	6.8	1868	1	US-08-069-863-1	Sequence 1, Appli
16	271.5	6.8	1868	1	US-08-309-069-1	Sequence 1, Appli
17	271.5	6.8	1868	1	US-08-562-833-1	Sequence 1, Appli
18	271.5	6.8	1868	4	US-09-659-749-1	Sequence 1, Appli
19	271.5	6.8	1868	4	US-09-672-105-1	Sequence 1, Appli
20	271.5	6.8	1868	5	PCT-US95-05520-1	Sequence 1, Appli
21	265.5	6.6	1146	2	US-08-504-265B-74	Sequence 74, Appli
22	264.5	6.6	8119	1	US-08-460-343B-1	Sequence 1, Appli
23	264.5	6.6	8119	1	US-08-398-028B-1	Sequence 1, Appli
24	264.5	6.6	8119	2	US-08-504-265B-1	Sequence 4, Appli
25	263.5	6.6	1140	3	US-09-023-173-4	Sequence 5, Appli
26	261	6.5	1330	4	US-09-966-921A-5	Sequence 6, Appli
27	260.5	6.5	1497	1	US-08-322-677A-6	Sequence 6, Appli
28	260.5	6.5	1497	3	US-08-898-218-6	Sequence 6, Appli
29	260.5	6.5	1497	3	US-08-848-793-6	Sequence 6, Appli
30	260.5	6.5	1497	4	US-09-445-270-1	Sequence 1, Appli
31	260.5	6.5	1497	4	US-09-178-173A-1	Sequence 1, Appli
32	260.5	6.5	1497	4	US-08-322-678-6	Sequence 6, Appli
33	260.5	6.5	1513	4	US-09-178-155-1	Sequence 1, Appli
34	260.5	6.5	1513	4	US-09-966-921A-1	Sequence 1, Appli
35	259	6.4	1306	4	RE34606-1	Patent No. RE34,60
36	258.5	6.4	1496	6	US-09-486-192-1	Sequence 1, Appli
37	257	6.4	2760	4	US-08-322-965-1	Sequence 1, Appli
38	253	6.3	1500	1	US-08-875-154-1	Sequence 1, Appli
39	252.5	6.3	10216	2	US-08-322-677A-11	Sequence 11, Appli
40	248	6.2	1140	1	US-08-322-676-11	Sequence 11, Appli
41	248	6.2	1140	1	US-08-898-218-11	Sequence 11, Appli
42	248	6.2	1140	3	US-08-848-793-11	Sequence 11, Appli
43	248	6.2	1140	4	US-09-445-270-6	Sequence 6, Appli
44	248	6.2	1140	4	US-08-322-678-11	Sequence 11, Appli
45	248	6.2	1140	4	US-08-322-678-11	Sequence 11, Appli

ALIGNMENTS

RESULT 1

US-07-671-376C-4
; Sequence 4, Application US/07671376C
; Patent No. 5294542
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan et al.
; TITLE OF INVENTION: RESIDUAL PROTEASE-III
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/671,376C
; FILING DATE: 19910319
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00811/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2532 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-07-671-376C-4

Alignment Scores:

Pred. No.: 4,51e-33 Length: 2532
 Score: 430.00 Matches: 198
 Percent Similarity: 39.13% Conservative: 135
 Best Local Similarity: 23.27% Mismatches: 302
 Query Match: 10.70% Indels: 216
 DB: 1 Gaps: 35

US-09-806-767-2 (1-775) x US-07-671-376C-4 (1-2532)

Qy 7 PheLeuCysIlellePheLeuPheCysSerSerSerGluIleLeu----- 23
 Db 61 TTTCTGCTTGAAGTTTCGCTTATTTTTCGGTTATCCACAGGCATTACGGCGGTTGAG 120
 Qy 23 ----- 23
 Db 121 GCAGCTCGGCTTCTTCAAAAAGCTCGCTGATCTGGAAAAAGCCGAGGTATTTCGGTGAT 180
 Qy 24 ----- 36
 Db 181 ATCGATATGACGACAAAGCAAAACACCGTTATAGTGGAAATAAAGAAAAATCCTTG 240
 Qy 37 -----GluThrAlaLysThr 41
 Db 241 CGAAGCGGAGGAGGAGGAGGAGCAATCGAAGCAAGCAAGCTGAAACCGCTCGCACC 300
 Qy 42 PheAlaSerLysPheAspTrpHisLeuSerPheLeuGlnGluAlaValGlu 61
 Db 301 AAAGCAAAA-----AACAAAGCAATCAAGCAGTGAGAA 333
 Qy 62 GluGluGluGluGluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGly 81
 Db 334 -----AACGGAAAAAGTAAACCGGGAATATGAGCAGGTATTCTCAGGC 375
 Qy 82 PheAlaAlaGlnLeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValVal 101
 Db 376 TTCTCTATGAAGCTTCAGCTATGAGATTCCAAAACTTTCAGCGTTAAAGACGTTAAG 435
 Qy 102 AlaValArgPro-----AspHisValLeuGlnValGlnThrThrTyr 115
 Db 436 CGAGTGACCCGACGTCACATATAAACACAGACAAATATGAGGATAAAGACGTCACAATC 495
 Qy 116 SerTyrLysPheLeuGly-----LeuAspGly-----PheGlyAsnSerGly 129
 Db 496 TCCGAAGACGCGGTATCTCCGCAAAATGGATGACAGTGGCGCTTATATCGGAGCAACGAT 555
 Qy 130 ValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyVal 149
 Db 556 GCATGGGATTTAGGCTACACAGGAAAGGCAATCAAGGTGGCGATTATTGACACTGGGGTT 615
 Qy 150 TrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLys 169
 Db 616 GAATACAAATCACCAGATCTGAAG-----AAA 642
 Qy 170 GlyIleCysGln-----GluGlyGluSerPheSerSerSerCysAsnArgLysLeuIle 188
 Db 643 AACTTTGGACAATATAAGGATACGATTTTGTGGACAATGATTACGAT----- 690
 Qy 189 GlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerPro--- 207
 Db 691 -----CCAAAGAAAGAACACCAACC 708
 Qy 208 AsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAla 227
 Db 709 GCGCATCCGAGG-----GGCAGGCAACTGACCATGGCAGACACAGTACGCC 753

Qy 228 SerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyVal 247
 Db 754 GGAACCTGTGGCTGCAAC-----GGAAACG 777
 Qy 248 AlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGly 267
 Db 778 ATTAAGCGGTAGCGCTGATCCACACTTCTTCCTTATCTGTTATCTGTTAGGCGCTGGCGGA 837
 Qy 268 CysTyrSerSerAsp---IleLeuAlaAlaIleAspValAlaIleGlnAspLysValAsp 286
 Db 838 AGCGGCACACGGAACAGTCATCGCGGGGTGGAACCTGCGAGTCGAGGACGGGGCAGAT 897
 Qy 287 ValLeuSerLeuSerLeuGly---GlyPheProIleProLeuTyrAspThrIleAla 305
 Db 898 GTGATGAACCTGTCTCTCGGAACCTCTTAAACAACCCGGACTGGGCGACACACAGCG 957
 Qy 306 IleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsn 325
 Db 958 CTTGACTGG-----GCCATGTCAAGAGGGGTTGTCGTCTGTACTCTCAACGCGCAACAGC 1011
 Qy 326 GlyProIleGluSerSerValAlaAsn-----ThrAlaProTrpValSerThrIleGly 343
 Db 1012 GGACCGAACGGCTGGACAGTCGGATCGCGGCGCACATCAAGAGAGCGATTCTTCGTCGGT 1071
 Qy 344 AlaGlyThrLeuAspArgPheProAlaValVal-----ArgLeuAlaAsnGlyLys 361
 Db 1072 GCGACTAGCTGCGCTCAATAGTACCGCTCACTTCGCTCCTACTCTTCAGCAAAA 1131
 Qy 362 LeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluVal 381
 Db 1132 GTGATGGGCTACAAACAAGAGGACGAGCTCAAGCGCTCAATAAC-----AAAGAAGTT 1185
 Qy 382 GluValIleTyrValThrGlyAspLysGlySerGluPheCysLeuArgGlySerLeu 401
 Db 1186 GAGCTTGTGTC---GAAGCGGGAATCGCGCAAGCAAGGATTTT----- 1224
 Qy 402 ProArgGluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSer 421
 Db 1225 GAAGGGAAGACCTTGACAGCAAGTCCCGTTGTCAACAGGACGAGCATTCGATTGTTGTG 1284
 Qy 422 GluLysGlyGluAlaValValLysGluAlaGlyValAlaMetIleLeuAlaAsnThrGlu 441
 Db 1285 GATAAAGCGGATAACGCTAAAGACCGGTGCAATCGCATGGTGTGTATTAACACCTC 1344
 Qy 442 IleAsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyr 461
 Db 1345 TCTGAGAAATGAAGCCATGTGCGAGCATGTCTGCCACAGCATTAAGCTTTCATTA 1404
 Qy 462 ThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIle 481
 Db 1405 GAAGACGCGCAAAACTC-----GTACGCGCTGAAAGCTGTGTGAGACAAAACA 1455
 Qy 482 IlePheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAla 501
 Db 1456 ACATTCAAGTTGACCGGTC---TCAAGACCGCTCGGTGACCAAGTCGCTGATTTCTATCA 1512
 Qy 502 ArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyVal 521
 Db 1513 CCGCGCCCT---GTTATGATACGTGGATGATTAAAGCTGATATTTCGCGCCAGGGGTC 1569
 Qy 522 AsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArg 541
 Db 1570 AATATCGTGAGCAGCATCCCAACACACGATCCTGCATACGCTATTCATTTGCCGAGCG 1629
 Qy 542 ArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIle 561
 Db 1630 CAA-----GGAAACAGCATGGCATCGCTTCATATTTCGCGAGCG 1668
 Qy 562 ThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeu 581
 Db 1669 GTTGGCGTTATTAAACAGCCCAACCAAGTGGAGCGTTGAACAGATTAAAGCCCATC 1728
 Qy 582 MetThrThrAla---AspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLys 600

```
Db 1729 ATGAATACCGCTGTCACCTTTAAAGGATAGCGATGGGGAAGTATATATCGCATAC-----1782
Qy 601 ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnPro 620
Db 1783 -----GTCAGGCGCGAGCGAGCGCAAGAAATTAAGACGCAATCAAGGCC 1827
Qy 621 GlyLeuValTyAsnIleGlnProValAspTyIleThrTyLeuCysThrLeuGly----639
Db 1828 GATTGCTGCTCCTACCTCGGAGCTATCATACGGCACGTTCTTTGAAGGAAACGGAAC 1887
Qy 640 PheThrArgSerAspIleLeuAlaIleThrHisLysAsn-----652
Db 1888 GAAACAAAATGAAGCTTTACGATTCAAAATCAATCTCCATTAGAAAGTCATACACA 1947
Qy 653 -----ValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeu-----667
Db 1948 CTTGAATACTCATTTAATGCGAGCGCATTTCCACATCCGCGCACAGCCGTGTTGTGATT 2007
Qy 667 -----667
Db 2008 CCGGCACATCAACCGGGAAGCCACTGCAAAAGCTAAAGTCAATACGAAGAAACAATA 2067
Qy 668 -----AsnTyProSerIleAlaValIlePheLysArgGlyLysThr-----681
Db 2068 GCTGGCACCATTAGAGGACGGTTATCGTCAGAGAGCGCGAAACGGTCGCTAAGGTA 2127
Qy 682 ---ThrGluMetIleThrArg-----ArgValThrAsnVal-----692
Db 2128 CCTACATTGCTGATTGTGAAAGAGCGCATTTATCCGAGAGTCACATCTGTCCTGTCAGC 2187
Qy 693 ---GlySerProAsnSerIleTySerValAsnValLysAlaProGluGly-----708
Db 2188 GAAGGGCTGTACAAAGTACTATCAAAATGAAACCTACCTTCTCGCGGAGCGGAAGAG 2247
Qy 709 IleLysValIleValAsnProLysArgLeu-----ValPheLys 721
Db 2248 CTGGCGTCTCTGCTCTATGACAGCAACCTTGATTTCGACGCCAAGCCGCAATTTATATA 2307
Qy 722 HisValAspGlnThrLeuSerTyArgValTrp 732
Db 2308 AACCAAGATAAAGGTTACCAGTACTTTGACTGG 2340
```

RESULT 2

```
US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-6
```

Alignment Scores:

```
Pred. No.: 1.14e-24 Length: 1977
Score: 345.00 Matches: 145
Percent Similarity: 33.17% Conservative: 62
Best Local Similarity: 23.24% Mismatches: 145
Query Match: 8.59% Indels: 272
DB: 3 Gaps: 21
```

US-09-806-767-2 (1-775) x US-08-894-818B-6 (1-1977)

```
Qy 29 lleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrp 48
Db 148 ATTCAAAATTTGAATCTTAACGAGAAATCAGACAGTAATTGTA-----TTTGAATAAC 201
Qy 49 HisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSer 68
Db 202 CAT-----ACGGAATAAAGAAATTGCA 222
Qy 69 SerArgLeuLeuTySerTyGlySer-----AlaIleGluGly 81
Db 223 GTACAGTCTCTTGAGTTAATCGGTGCAAAAGTAGGTATGTACCATATTATACCCGCA 282
Qy 82 PheAlaAlaGln-----LeuThrGluSerGlu 90
Db 283 ATAGCTGCCGATCTTAAGGTTAGACACTTACTAGTCATCTCAGGTTTAAACAGGGGTAAA 342
Qy 91 AlaGluIle-----LeuArgTySerProGluValValAlaValArg-----104
Db 343 GCTAAGCTTTTCAGGTGTAGGTTTATCCAGGAAGACTACAAAGTTACAGTTTCACAGAA 402
Qy 105 -----ProAspHisValLeuGlnValGlnThrThrTySerTyLysPheLeu 120
Db 403 TTAGAAGACTGGATGAGTCTGCAGCTCAAGTTATGGCACTTAC-----447
Qy 121 GlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThr 140
Db 448 -----GTTTGGAACTTGGGATATGATGTTCTGGAATC 480
Qy 141 lleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspThrGly 160
Db 481 ACAATAGGAATAAATTGACACTGGAATT-----GACGCTTCT 516
Qy 161 MetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSer 180
Db 517 CATCCAGATCTC-----528
Qy 181 SerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAla 200
Db 529 -----CAAGGAAGTAATTGGGTGGGTAGATTGTTGTCATGGT-----567
Qy 201 AsnSerProGluGluSerProAsnMetProArgGluTyIleSerAlaArgAspSerThr 220
```

Db 568 -----AGAGTTAT-----CCATACGATGACCAT 591
Qy 221 GlyHisLeuThrHisThrAlaSerThrValGlyGlySerValSerMetAlaVal 240
Db 592 GGACATGGAACTCATGTAGCTTCAATAGCAGCTGTACTGGACGACGA 639
Qy 241 LeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisLeuAlaValTyr 260
Db 640 -----AGTAATGGCAAGTAGCAAGGAATGGCTCCAGGAGCTAAGCTGGCGGAAT 690
Qy 261 LysValCysTrpPheAsnGlyCysTyrSer---SerAspIleLeuAlaAlaIleVal 279
Db 691 AAGTTCTAGTCCGATGCTTCTGGAAGCATATCTACTATTAATTAAGGAGTTGAGTGG 750
Qy 280 AlaIleGlnAspLys-----ValAspValLeuSerLeuSerLeuGlyGlyPhe 295
Db 751 GCCGTTGATACAAAGATAGTAGCAAGTAATTAAGTTCATTAATCTTCTCTGTTCAAGC 810
Qy 296 ProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGly 315
Db 811 CAGAGCTCCGACGGAACCGACTCCCTCAGTCAGCGCGTCAACAAACGCTGGACGCGGT 870
Qy 316 IleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThr 335
Db 871 ATAGTAGTCTGCTGCGCGCGCAACAGCGGCGCAACACCTACACCGCTCGCTCACCC 930
Qy 336 AlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaVal 355
Db 931 GCCCGCGGACGAGGTATACCGTCGTCGTCAGTTGACAGCAAC 975
Qy 356 ArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLys 375
Db 975 ----- 975
Qy 376 AsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGlySerGluPhe 395
Db 975 ----- 975
Qy 396 CysLeuArgGlySerLeuProAGluGluIleArgGlyLysMetValIleCysAspArg 415
Db 975 ----- 975
Qy 436 IleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeuLeuPro 455
Db 975 ----- 975
Qy 456 AlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrVal 475
Db 975 ----- 975
Qy 476 LysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
Db 976 -----GACAC 981
Qy 496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515
Db 982 ATCCGACGTTCTCCAGCAGGGGACCGCGC-----GACGGAAGCTCAAGCGGAA 1035
Qy 516 MetIleAlaProGlyValAsnIleAlaAlaTrpProGlnAsnLeuGlyProThr--- 534
Db 1036 GTCGTCGCGCGCGCGTGTACATCATAGCC-----CCGCGCGCCAGCGGAACCATG 1089
Qy 535 GlyLeuProTyrAspSerArgValAlaAsnPheThrValMetSerGlyThrSerMetSer 554
Db 1090 GGCACCCCGATACAGAC-----TACTACCAAGCGCTCTGGAACCATGAGGCC 1140
Qy 555 CysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerPro 574

Db 1141 ACCCGCACGTTTGGCGGTTGGCGGCTCATCTCCAGGCCACCGAGCTGACCCCG 1200
Qy 575 AlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAla 594
Db 1201 GACAAGGTGAAGACCGCCCTCATCGAGACCGCCGACATATGTC-----GCCCCCAAGGAG 1254
Qy 595 IleLysAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro 614
Db 1255 ATACCGGAC-----ATCGCCTACGTCGGGTAGGTTGAACGTC 1293
Qy 615 GlnLysAlaIle 618
Db 1294 TACAAGGCCATC 1305
RESULT 3
US-08-894-818B-34
; Sequence 34, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikumoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1962 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-894-818B-34
Alignment Scores:
Pred. No.: 2,32e-23 Length: 1962
Score: 332.00 Matches: 140
Percent Similarity: 22.75% Conservative: 65
Best Local Similarity: 22.36% Mismatches: 147
Query Match: 8.26% Indels: 274
DB: 3 Gaps: 21

US-09-806-767-2 (1-775) x US-08-894-818B-34 (1-1962)

QY 29 IleValGlnLeuHisProAsnSerGluThrAlaIleValThrPheAlaSerLysPheAspTrp 48
Db 148 ATTCAAAATTTGAATCTTACGAGGAATACAGACACAGTAATTGTA-----TTTGAAC 201
QY 49 HisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluGluProSer 68
Db 202 CAT-----AGGGAAGAAAGAAATTGCA 222
QY 69 SerArgLeuLeuTySerTyGlySer-----AlaIleGluGly 81
Db 223 GTAAGAGTTCTTGAGTTAATGGTGCAAAAGTTAGGTATGTACCATATATTATACCGCA 282
QY 82 PheAlaAlaGln-----LeuThrGluSerGlu 90
Db 283 ATAGCTCCGATCTTAAGTTAGAGACTTACTAGTCATCTCAGGTTTAAACAGGGGTAAA 342
QY 91 AlaGluIle-----LeuArgTySerProGluValValAlaValArg----- 104
Db 343 GCTAAGCTTTCAGGTGTAGTTTATCCAGGAAGACTACAAGTTACAGTTTCAGCAGAA 402
QY 105 -----ProAspHisValLeuGlnValGlnThrThrTySerTyLysPheLeu 120
Db 403 TTAGAGGACTGGATGAGTCTGCAGCTCAAGTTATGCACTTAC----- 447
QY 121 GlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThr 140
Db 448 -----GTTTGAACCTGGGATATGATGTTCTGGAATC 480
QY 141 IleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspAspThrGly 160
Db 481 ACAATAGGAATAATGACACTGGAATT-----GACGTTCT 516
QY 161 MetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSer 180
Db 517 CATCCAGATCTC----- 528
QY 181 SerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAla 200
Db 529 -----CAAGGAAAAGTAATTTGGTGGTAGATTTTGTCAATGTT----- 567
QY 201 AsnSerProGluGluSerProAsnMetProArgGluTyIleSerAlaArgAspSerThr 220
Db 568 -----AGGAGTTAT-----CCATACGATGACCAT 591
QY 221 GlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnVal 240
Db 592 GGACATGGAATCATATGATGCTCAATAGCTGCTACTAGCAGCTGCTACTGAGCAGCA----- 639
QY 241 LeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTy 260
Db 640 -----AGTAATGGCAAGTCAAGGAATGCTCCAGGAGCTAGCTGGCGGNATT 690
QY 261 LysValCysTrpPheAsnGlyCysTySer-----SerAspIleLeuAlaIleAspVal 279
Db 691 AAGGTTCTAGGTGCGGTGCTGCTGGAAGCATATCTACTATAATTAAGGAGTTGAGTGG 750
QY 280 AlaIleGlnAspLys-----ValAspValLeuSerLeuSerLeuGlyGlyPhe 295
Db 751 GCCGTTGTAACAAAGATAAGTACGGAATTAAAGTTCATTATCTTCTCTGTTGTTCAAGC 810
QY 296 ProIleProLeuTyAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGly 315
Db 811 CAGAGCTCAGATGTTGATGCGCTCTTAAGTCAGGCTGTTAATGAGCGCTGGATGCTGGA 870
QY 316 IleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThr 335
Db 871 TTACTGTTGTTGTTGCTGCTGGAACAGTGCAGCTAACAAGTATACATCGGTTCTCCA 930
QY 336 AlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValVal 355

RESULT 4

US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6356726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997

```

; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15

```

Alignment Scores:

Pred. No.:	2,328-23	Length:	1962
Score:	332.00	Matches:	140
Percent Similarity:	32.75%	Conservative:	65
Best Local Similarity:	22.36%	Mismatches:	147
Query Match:	8.26%	Indels:	274
DB:	4	Gaps:	21

US-09-806-767-2 (1-775) x US-09-445-472-15 (1-1962)

Qy	29	IleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerIysPheAspTrp	48
Db	148	ATTCAAAATTCGAATCCTAACGAGGAATCAGCACAGTAATGTGA-----TTTGAATAAC	201
Qy	49	HisLeuSerPheLeuGlnCluAlaValLeuGlyValGluGluGluGluGluGluGluProSer	68
Db	202	CAT-----AGGGAAAAAGAATAATTGCA 222	
Qy	69	SerArgLeuLeuTyxSerTyxClySer-----AlaIleGluGly 81	
Db	223	GTAAGAGTCTTCCTGAGTTAATGGTGCCAAAAGTAGGTATGTACCATTATATACCGCA 282	
Qy	82	PheAlaAlaGln-----LeuThrGluSerGlu 90	
Db	283	ATAGCTGCAGACTTAAAGGTTAGAGACTTACTAGTCATCTCAGGTTTACAGGGGGTAA 342	
Qy	91	AlaGluIle-----LeuArgTyrSerProGluValValAlaValArg----- 104	
Db	343	GCTAAGCTTTCAGGTGTAGGTATTATCCAGAGNACTACAAGTTACAGTTTACAGCAA 402	
Qy	105	-----ProAspHisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeu 120	
Db	403	TTAGAGGACTGGATGAGTCTCAGCTCAAGTTATGGCACTTAC----- 447	
Qy	121	GlyLeuAspGlyPheGlyAsnSerGlyValTrpSerIysSerArgPheGlyGlnGlyThr 140	
Db	448	-----GTTTGGAACTGGCATATGATGGTCTTCGGAATC 480	
Qy	141	IleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspAspThrGly 160	
Db	481	ACAATAGGAATAATTGACACTCGAATT-----GACGTTCT 516	
Qy	161	MetProSerIleProArgIysTrpIysGlyIleCysGlnGluGlyGluSerPheSerSer 180	
Db	517	CATCCAGATCTC----- 528	
Qy	181	SerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAla 200	
Db	529	-----CAAGAAAAGTAATTGGTGGGTAGATTTTGTCAATGTT----- 567	
Qy	201	AsnSerProGluGluSerProAsnMetProArgGluTyrIleSerAlaAlaArgAspSerThr 220	
Db	568	-----AGGAGTTAT-----CCATACGATGACCAT 591	
Qy	221	GlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnVal 240	
Db	592	GGACATGGAACTCATCTAGCTTCATAGCAGCTGCTACTGGAGCAGCA----- 639	
Qy	241	LeuGlyAsnGlyAlaGlyValAlaArgGlyMetalapProGlyAlaHisIleAlaValTyr 260	
Db	640	-----AGTAATGGCAAGTACAGGAATATGGTCCAGGAGCTTAAGCTGCCGGGAATT 690	

261	Qy	LysValCysTrpPheAsnGlyCysTrpSer	---SerAspIleLeuAlaIleAspVal	279
691	Db	AAAGTTCTAGTCCGATGGTTCTGGAAGCATATCTACATATAATTAAGGAGTGTGAGTGG	750	
280	Qy	AlaIleGlnAspLys	-----ValAspValLeuSerLeuSerLeuGlyGlyPhe	295
751	Db	GCCTTGATTAACAAGATAGTACGGAATTAAGTTCATTAATCTTCTCTGTGGTCAAGC	810	
296	Qy	ProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGly	315	
811	Db	CAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGACGCGTGGGATGCTGGA	870	
316	Qy	IleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThr	335	
871	Db	TTAGTTTCTGGTTCGGCTCGAACAACAGTGGACCTAACAAAGATATACAACTGGTCTCTCCA	930	
336	Qy	AlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValVal	355	
931	Db	GCAGCTCAAGCAAGTTATTACAGTTGGAGCGGTTGACAAG	972	
356	Qy	ArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLys	375	
972	Db	-----	972	
376	Qy	AsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGlySerGluPhe	395	
972	Db	-----	972	
396	Qy	CysIleuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArg	415	
972	Db	-----	972	
416	Qy	GlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMet	435	
972	Db	-----	972	
436	Qy	IleLeuAlaAsnThrGluIleAsnGlnGluLysAspSerIleAspValHisLeuLeuPro	455	
973	Db	-----TATGATGTT-----	981	
456	Qy	AlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrVal	475	
981	Db	-----	981	
476	Qy	LysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu	495	
981	Db	-----	981	
496	Qy	ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp	515	
982	Db	ATAACAAGCTTCTCAAGCAGAGGCGCAACTGCA-----GACGCGAGCTTTAAGCCTGAG	1035	
516	Qy	MetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGly	535	
1036	Db	GTGTGTGCTCCAGAAACTGGATTAATGTCTCCAGACCAAGTGGAACTAGC---ATGGGT	1092	
536	Qy	LeuProTyrAspSerArgValAsnPheThrValMetSerGlyThrSerMetSerCys	555	
1093	Db	CAACCAATTATGAC-----TATTACACAGCAGCTCTCGGACATCAATGGCACT	1143	
556	Qy	ProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAla	575	
1144	Db	CCTCAGCTAGTGGTATTGTCAGCCCTCTTGCTCCAAGCACACCGAGCTGACCTCCAGAC	1203	
576	Qy	AlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIle	595	
1204	Db	AAAGTAAACACAGCCCTCATAGAACGTCTGATATCGTA-----	1242	
596	Qy	LysAspGlyAsnLysProAlaGlyVal-----PheAlaIleGlyAlaGlyHisValAsn	613	
1243	Db	-----AAGCCAGATGAATAGCCGATATAGCTACGGTGCAGGTTAGGTTAAT	1290	
614	Qy	ProGlnLysAlaIleAsn	619	

```

Db      1291 GCATACAGGCTATAAC 1308
      |||||||
RESULT 5
US-08-894-818B-2
; Sequence 2, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-894-818B-2
Alignment Scores:
Pred. No.: 3,74e-23 Length: 1977
Score: 330.00 Matches: 124
Percent Similarity: 34.54% Conservative: 57
Best Local Similarity: 23.66% Mismatches: 119
Query Match: 8.21% Indels: 224
DB: 3 Gaps: 17

US-09-806-767-2 (1-775) x US-08-894-818B-2 (1-1977)
QY      103 ValArgProAspHisValLeuGlnThrThrTyrSerTyrLysPheLeuGlyLeu 122
Db      382 ATACGAGGATTACAAGGTTGAGTGGACGCCACTTCGGTCTCCAGATAGGGGCC 441
QY      123 AspGlyPheGlyAsnSerGlyValTyrSerTyrSerArgPhe---GlyGlnGlyThrIle 141
Db      442 GAT-----ACCGTCTGGAACCTCCCTCGCTACGACGGAACGGGTGGTG 486

142 IleGlyValLeuAspThrGlyValTyrProGluSerProSerPheAspThrGlyMet 161
Db      487 GTTGCATCGTGCATACGGGTATA-----GACGGGAACCAAC 522
QY      162 ProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSer 181
Db      523 CCCGATCTG-----AAGGEC----- 537
QY      182 SerCysAsnArgLysLeuIleGlyAlaAArgPheIleArgGlyHisArgValAlaAsn 201
Db      538 -----AAGTTCATAGGCTGTACGACGCCGTCAACGGCAGGTGCG----- 576
QY      202 SerProGluGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGly 221
Db      577 -----ACCCCTACGATGACACGAGGA 597
QY      222 HisGlyThrHisThrAlaSerThrValGlyGlySer---SerValSerMetAlaAsnVal 240
Db      598 CACGGAACCCACGTTGCGGGTATCTTCCCGAAGCCGCGAGCGTAACTCCAGTACATA 657
QY      241 LeuGlyAsnGlyValAlaAArgGlyMetAlaProGlyAlaHisIleAlaValTyr 260
Db      658 -----GGCGTGGCCCCGGCGGAAGCTCGTCGGCGTC 690
QY      261 LysValCysTrpPheAsnGlyCysTyrSer---SerAspIleLeuAlaAlaAspVal 279
Db      691 AAGGTTCTCGGTGCGGACGGTTCGGGAAGCGTCTCCACCATCATCGCGGGTGTGACTGG 750
QY      280 AlaIleGlnAspLys-----ValAspValLeuSerLeuSerLeuGlyGlyPhe 295
Db      751 GTCGTCCAGAAACAGGACAAGTACGGGATAAGGGTTCATCAACCTCTCCCTCGGTCCTCC 810
QY      296 ProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGly 315
Db      811 CAGAGCTCGACGGAACCGACTCCCTCAGTCAGGCCGTCAACACCGCTGGGACGCCGGT 870
QY      316 IleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThr 335
Db      871 ATAGTAGTCTCGTCTCGCGCGCGCAACAGCGGGCGGAACACTACACCGTGGCTCACCC 930
QY      336 AlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValVal 355
Db      931 GCCCGCGGACGACAGGTCTATAACCGTCGTCGTCGAGTTCACAGCAAC----- 975
QY      356 ArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLys 375
Db      975 ----- 975
QY      376 AsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGlySerGluPhe 395
Db      975 ----- 975
QY      396 CysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArg 415
Db      975 ----- 975
QY      416 GlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyValAlaMet 435
Db      975 ----- 975
QY      436 IleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeuLeuPro 455
Db      975 ----- 975
QY      456 AlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrVal 475
Db      975 ----- 975
QY      476 LysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
Db      976 -----GACAAC 981
QY      496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515

```


Qy	516	MetIleAlaProGlyValAsnIleAlaAlaTrrProGlnAsnLeuGlyProThr---	534
		::::: ::::: ::::: ::::: ::::: :::::	
Db	1036	GTCTGCCCCCGCGCTTGATCATAGCC-----CCGCGCGCCAGCGAACCCAGCATG	1089
Qy	535	GlyLeuProTrrYrAspSerArgValAsnPhetrrValMetSerGlyThrSerMetSer	554
		::::: ::::: ::::: ::::: ::::: :::::	
Db	1090	GGCACCCCGATACAGAC-----TACTACCAAGGCCTCTGGAACGAGCATGGCC	1140
Qy	555	CysProHisValSerGlyIleThrAlaLeuIleArgSerAlatYrProAsnTrrSerPro	574
		::::: ::::: ::::: ::::: ::::: :::::	
Db	1141	ACCCCGCACGTTTCGGCGGTGGCGGCTCATCTCCAGGCCACCACCGACTGCACCCCG	1200
Qy	575	AlaalaIleYsSerAlaLeuMetThrThralaAspLeuYrAspArgGlnGlyLysAla	594
		::::: ::::: ::::: ::::: ::::: :::::	
Db	1201	GACAAGGTGAAGACCGCCCTCATCGACCGCCGACATAGTC-----GCCCCCAAGGAG	1254
Qy	595	IleYsAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro	614
		::::: ::::: ::::: ::::: ::::: :::::	
Db	1255	ATACCGAC-----ATCGCCTACGGTCGGGTAGGTGAACGTC	1293
Qy	615	GlnLysAlaIle	618
Db	1294	TACAAGGCCCATC	1305

RESULT 7

```

US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tonoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2

```

```

Alignment Scores:
Pred. No.:      5,978-23      Length:      1236
Score:          324.50      Matches:      119
Percent Similarity: 32.88%      Conservative: 51
Best Local Similarity: 23.02%      Mismatches:  116
Query Match:      8.08%      Indels:      231
DB:              4          Gaps:         15

US-09-806-767-2 (1-775) x US-09-445-472-2 (1-1236)

Qy      110  GlnValGlnThrTyrSerTyrLysPheLeuGlyLeuAspGlyPheGlyAsnSerGly 129
      |||||
Db      34  CAAGTTATGGCAACTTAC----- 51

Qy      130  ValTrrSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyVal 149
      |||||
Db      52  GTTTGGAACTTGGGTATGATGGTTCTGGATACACATAGGATTAATGCATCGAATT 111

Qy      150  TrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLys 169
      |||||
Db      112 -----GACGCTTCTCATCCAGATCTC----- 132

```

Qy	170	GlyIleCysGlnGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIleGly	189
Db	133	-----CAAGGAAAGTAAATGGG	150
Qy	190	AlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMet	209
Db	151	-----TGGGTAGATTTGTCAATGGT	171
Qy	210	ProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThr	229
Db	172	--AGGAGTTAT-----CCATACGATGACCATGCATGCACTCATGTAGTTCATAA	222
Qy	230	ValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArg	249
Db	223	GCAGCTGGTACTGAGCAGCA-----AGTAATGGCAAGTACAAAG	261
Qy	250	GlyMetAlaProGlyValAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTyr	269
Db	262	GGAATGGTCCAGAGCTAGCTGGCGGAATTAAAGTTCTAGGTGCGGATGGTCTCGGA	321
Qy	270	Ser---SerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLys-----	284
Db	322	AGCATATCTACTATAATTAAAGGAGTTGAGTGGCCGTTGATAACAAAGATAAGTACGGA	381
Qy	285	ValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIle	304
Db	382	ATTAAAGTCTAATAATCTTCTCTGGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTA	441
Qy	305	AlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaIleGlyAsn	324
Db	442	AGTCAGCGTGTAAATGACGCGTGGAGTCTCGATAGTTGTTGGTTCGCGCTCGAAGC	501
Qy	325	AsnGlyProIleGluSerSerValAlaAsnThrAlaProTyrValSerThrIleGlyAla	344
Db	502	AGTGGACCTAAAGATATACATCAATCGGTCTCCAGCAGCTCGCAAGAAAGTTATTACAGT	561
Qy	345	GlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyr	364
Db	562	GGAGCGGTTGCAAG-----	576
Qy	365	GlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIle	384
Db	576	-----	576
Qy	385	TyrValThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGlu	404
Db	576	-----	576
Qy	405	GluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGly	424
Db	576	-----	576
Qy	425	GluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGln	444
Db	576	-----	576
Qy	445	GluGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSer	464
Db	577	-----TATGATGTT-----	585
Qy	465	ValLeuLeuLysAlaTyrValAlaAsnAlaThrValLysProLysAlaArgIleIlePheGly	484
Db	585	-----	585
Qy	485	GlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyPro	504
Db	586	-----ATAACAAGCTTCTCAAGCAGACGAGCGCCA	612
Qy	505	SerLeuAlaAsnProSerIleIleLysProAspMetIleAlaProGlyValAsnIleIle	524
Db	613	ACTGCA-----GACGGCAGCGCTTAAGCCTGAGGTGTGTCTCCAGGAACATGGATAATT	566
Qy	525	AlaAlaTyrProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsn	544

Db	667	GCTGCCAGAGCAAGTGGAACTAGC-----ATGGGTCAACCAATTAATGAC-----TAT	714
Qy	545	PheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeu	564
Db	715	TACACAGCAGCTCCTGGGACATCAATGCAACTCTCAGTAGCTGGTATTGACGCCCTC	774
Qy	565	IleArgSerAlaTyProbsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThr	584
Db	775	TTGTCTCAAGCACACCCAGCTGGACTCCAGACAAAGTAAACACAGCCCTCATAGAACT	834
Qy	585	AlaaspLeuTyAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAlaGlyVal	604
Db	835	GCTGATATCGTA-----AAGCCAGATGAATA	861
Qy	605	-----PheAlaIleGlyAlaGlyHisValasnProGlnLysAlaIleAsn	619
Db	862	GCGGATATAGCTACGCTGCAGGTAGGTTATGTCATACAGGCTATTAAC	912
RESULT 8			
US-08-894-818B-4			
; Sequence 4, Application US/08894818B			
; Patent No. 6261822			
; GENERAL INFORMATION:			
; APPLICANT: TAKAKURA, Hikaru			
; APPLICANT: MORISHITA, Mio			
; APPLICANT: YAMAMOTO, Katsuhiko			
; APPLICANT: MITTA, Masanori			
; APPLICANT: ASADA, Kiyozo			
; APPLICANT: TSUNASAWA, Susumu			
; APPLICANT: KATO, Ikunoshin			
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES			
; NUMBER OF SEQUENCES: 42			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Browdy and Neimark			
; STREET: 419 Seventh Street N.W., Ste. 300			
; CITY: Washington			
; STATE: D.C.			
; COUNTRY: United States of America			
; ZIP: 20004			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/894,818B			
; FILING DATE: 20-MAY-1998			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: PCT/JP96/03253			
; FILING DATE: 07-NOV-1996			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: JP 323285/1995			
; FILING DATE: 12-DEC-1995			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Browdy, Roger L.			
; REGISTRATION NUMBER: 25,618			
; REFERENCE/DOCKET NUMBER: TAKAKURA-1			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (202) 628-5197			
; TELEFAX: (202) 737-3528			
; INFORMATION FOR SEQ ID NO: 4:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 1566 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: linear			
; MOLECULE TYPE: genomic DNA			
; FEATURE:			
; OTHER INFORMATION: /note= N at position 1283 is G or T.			
US-08-894-818B-4			

Alignment Scores:			
Pred. No.:	8.98e-23	Length:	1566
Score:	324.50	Matches:	119
Percent Similarity:	32.88%	Conservative:	51
Best Local Similarity:	23.02%	Mismatches:	116
Query Match:	8.08%	Indels:	231
DB:	3	Gaps:	15
US-09-806-767-2 (1-775) x US-08-894-818B-4 (1-1566)			
Qy	110	GlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGlyPheGlyAsnSerGly	129
Db	34	CAAGTTATGCAACTTAC-----	51
Qy	130	ValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyVal	149
Db	52	GTTTGGAACCTGGGATATGATGTTCTGGAATCACAATAGGAATAATGACACTGGAAT	111
Qy	150	TrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLys	169
Db	112	-----GACGCTTCTCATCCAGATCTC-----	132
Qy	170	GlyIleCysGlnGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIleGly	189
Db	133	-----CAAGGAAAAGTAATTTGGG	150
Qy	190	AlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMet	209
Db	151	TGGGTAGATTTTCTCAATGGT-----	171
Qy	210	ProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThr	229
Db	172	---AGGAGTTAT-----CATACGATGACCATGGACATGGAACCTCATGTAGTTCAATA	222
Qy	230	ValGlyGlySerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArg	249
Db	223	GCAGCTGGTACTGGAGCAGCA-----AGTAATGGCAAGTACAAG	261
Qy	250	GlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTyr	269
Db	262	GGAATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTGCGGATGTTCTGGA	321
Qy	270	Ser---SerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLys-----	284
Db	322	AGCATATCTACTAATAATTAAGGGAGTTGAGTGGCGCGTTGATACAAAGATAAGTACCGA	381
Qy	285	ValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIle	304
Db	382	ATTAAGTTCATTAATCTTTCTCTGGTTCAGCCAGAGCTCAGATGCTACTGACGCTCTA	441
Qy	305	AlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsn	324
Db	442	AGTCAGGCTGTTAATGCAGCGTGGGATGCTGGATTAGTTGTTGGTTGCGCTGGAAC	501
Qy	325	AsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyVala	344
Db	502	AGTGGACCTAAAGATATACATCGTTCTCCACAGCTCGCAAGCAAGTATTACAGTT	561
Qy	345	GlyThrLeuAspArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyr	364
Db	562	GGAGCGGTTGACAAG-----	576
Qy	365	GlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIle	384
Db	576	-----	576
Qy	385	TyrValThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGlu	404
Db	576	-----	576
Qy	405	GluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGly	424
Db	576	-----	576

QY 425 GluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGln 444
Db 576 -----
QY 445 GluGluAspSerIleAspValHisLeuLeuProAlaThrIleuIleGlyTyrThrGluSer 464
Db 577 -----TATGATGTT-----
QY 465 ValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGly 484
Db 585 -----
QY 485 GlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyPro 504
Db 586 -----ATAACAAGCTTCTCAAGCAGAGGCCA 612
QY 505 SerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAlaIleIle 524
Db 613 ACTGCA-----GACGGCAGGCTTAAGCTGAGGTTGTTCTCAGGAAACTGGATAATT 666
QY 525 AlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsn 544
Db 667 GCTCCACAGCAAGTGAAGTACG---ATGGGTCAACCAATTATGAC-----TAT 714
QY 545 PheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeu 564
Db 715 TACACAGCAGCTCTCTGGACATCAATGCGCACTCTCAGCTAGTGTATTCAGGCCCTC 774
QY 565 IleArgSerIleTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThr 584
Db 775 TTGCTCCAAGCACACCCGAGCTGACTCCAGACAAAGTAAACACAGCCCTCATAGAACT 834
QY 585 AlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAlaGlyVal 604
Db 835 GCTGATATCGTA-----AAGCCAGATGAATA 861
QY 605 -----PheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn 619
Db 862 GCCGATATAGCTACGTTGACAGGTAGGTTAATGATACACAGGCTATAAAC 912
RESULT 9
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 203 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLSCULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
US-09-000-016-3
Alignment Scores:
Pred. No.: 1,248-19 Length: 2539
Score: 297.00 Matches: 155
Percent Similarity: 31.03% Conservative: 70
Best Local Similarity: 21.38% Mismatches: 224
Query Match: 7.39% Indels: 276
DB: 3 Gaps: 22
US-09-806-767-2 (1-775) x US-09-000-016-3 (1-2539)
QY 36 SerGluThrAlaLysThrPheAlaSer---LysPheAspTrpHisLeuSerPheLeuGln 54
Db 632 GGGGACGGCGCGCGTGGTGGCGAGCGCGAGCTCGACCCGGCGCTCTTCGACATCACC 691
QY 55 GluAlaValLeuGlyValIleGluGluProSerSerArgLeuTyrSer 74
Db 692 GAATCGCGCAGCGCGCAGCGCACTCCAGAAACAGGAGCTGAAGTTCATCTCGCGC 751
QY 75 TyrGlySerAlaIleGluGlyPheAlaGln----- 85
Db 752 TACCAGGGCGCGCAGCGCGCGCCAGAGCGCGGCGGCGAACTCCGCCGG 811
QY 86 ---LeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArg 104
Db 812 ACCCTGAGCTCCCTGAACCGGACCGGTGGCGACCCCGCAGCAGCGCGTCCGAGCTG 871
QY 105 ProAspHisValLeuGlnValGlnThrThrTyrSer---TyrLysPheLeuGlyLeuAsp 123
Db 872 TGGGACGCGCTCACCAACGCGCAGCGGACCGCCCTCGGCATCGCCACGCTCTGGTGGAC 931
QY 124 Gly-----PheGlyAsnSerGlyValTrp 131
Db 932 GGGGTCCGACGGCGCGCTCGACACGCTCCGTGGGCGAGATCGGCGCCCAAGCGTGG 991
QY 132 SerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyValTrpPro 151
Db 992 TCCGCGCGCTACGACGCGAAGCGGTGAAGATCGCGCTCGACACCGGTGTC----- 1045
QY 152 GluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLysGlyIle 171
Db 1046 -----GACACGAGCATCGGACCTG-----AAGGCGCG 1075
QY 172 CysGlnGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIleGlyAlaArg 191
Db 1076 GTGACCGCGCTCCAGAACTTCACCGCGCGCGCC----- 1108
QY 192 PhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArg 211
Db 1108 ----- 1108
QY 212 GluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGly 231

Db 1109 -----GGCGCCGGCGACAGGTGGCGCACCGCACCGTCGCTCGATCGCGCG 1159
Qy 232 GlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMet 251
Db 1160 GGACAGCGGC-----GCCAGTCCAGGCGCAAGTACAAAGGCGGTC 1198
Qy 252 AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys--TyrSer 270
Db 1199 GCACCGCGCGCCCGGATCTCAACGGCAGAGTCTCGACGATCCGGTTTCGGCGACGAC 1258
Qy 271 SerAspIleLeuAlaIleAlaIleAlaIleGlnAspLysValAspValLeuSerLeu 290
Db 1259 TCCGGCATCTCCCGCATGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1318
Qy 291 SerLeuGlyGlyPheProIleTyrAspAspThrIleAlaIleGlyThrPheArg 310
Db 1319 AGCTGGCGCGCATGACACACCGGAGACCGACCGCTCGAGCGCGCGCGCGCGCGCGCG 1378
Qy 311 AlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnGlyProIleGluSer 330
Db 1379 TCCGCGGAGAGGCGGTCTGTTCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1438
Qy 331 SerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArg 350
Db 1439 GGTTCGCGCGCGAGCG 1480
Qy 351 PheProAlaValValArgLeuAlaAsnGlyLysLeuTyrGlyGluSerLeuTyrPro 370
Db 1480 ----- 1480
Qy 371 GlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAsp 390
Db 1481 -----GTCCAGCACAGGAC 1495
Qy 391 LysGlySerGluPheCysLeuArgGlySerLeuProArgGluIleArgGlyLysMet 410
Db 1496 AAG----- 1498
Qy 411 ValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAla 430
Db 1498 ----- 1498
Qy 431 GlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAsp 450
Db 1498 ----- 1498
Qy 451 ValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyr 470
Db 1498 ----- 1498
Qy 471 ValAsnAlaThrValLysProLysAlaArgIlePheGlyGlyThrValIleGlyArg 490
Db 1498 ----- 1498
Qy 491 SerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSer 510
Db 1499 -----CTCGCGGACTTCTCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1543
Qy 511 IleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaIleAlaIleProGln--- 529
Db 1544 ATC---AAGCGGAGCGTCACCGCTCCCGGCGTGACATCACCGCGCGCGCGCGCGCGCG 1600
Qy 530 -----AsnLeuGlyProThrGlyLeuProTyrAspSerArgArg 542
Db 1601 AACGACATCGCGCCAGGAGTGGTGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1639
Qy 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1640 -----TACATGACCATCTCCGGACATCGATGGCGACCGCGCGCGCGCGCGCGCGCG 1693
Qy 563 AlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaIleLysSerAlaLeuMet 582
Db 1694 GCCCTCTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1750

Qy 583 ThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db 1751 -----ACCGCTCCACCAAGGC 1768
Qy 603 Gly-----ValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn 619
Db 1769 GGCAAGTACACCGCGTTCGAGCAGGCTCGGGCGGATCCAGGCCGACACAGGGCGCTCCAG 1828
Qy 620 ProGlyLeuVal-----TyrAsnIleGln----- 627
Db 1829 CAGACCGTGTATCGCGCACCGCGTCTCGGTGAGTTCGGGTCCAGCAGTGGCGCGCACACC 1888
Qy 628 -----ProVal---AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSer 643
Db 1889 GACGACGAGCGGTCTACCAAGCAGCTGACCTACGCG---AACCTCGCG---ACCAGGAC 1942
Qy 644 AspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnPro 663
Db 1943 GTACAGCTGAAGCTGACGTGACGCGCACCGCACCGCAAGCGCGCGCGCGCGCGCGCGCG 2002
Qy 664 GlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrGlu 683
Db 2003 TTCTTTCAGCTGGCGCGCACCGCTGACGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 2062
Qy 684 MetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnVal 703
Db 2063 ATG---ACCGCGCACCGCGCTCGCGCGCACCGGTGACCGCGCGCTACTCGCGTACGTG 2119
Qy 704 LysAlaProGluGly 708
Db 2120 GTCCCGCAGGCGCGC 2134
RESULT 10
US-09-514-340-3
; Sequence 3, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIDINE D
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs

Db	1199	GCACCGGGCGCGGATCTCAAGCGCAAGGTCTCTGACGACTCCGGTTTCGGCGACGAC	1358
Qy	271	SerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeu	290
Db	1259	TCCGGCATCTCCCGCATGGAGTGGCGCGCGCGAGCGCGACGTCGTCAACCATG	1318
Qy	291	SerLeuGlyGlyPheProIleProLeuTyrrAspAspThrIleAlaIleGlyThrPheArg	310
Db	1319	AGCCTGGCGGCATGACACACCGAGACCGACCCCTGGAGCGCGGTCGACCAAGCTG	1378
Qy	311	AlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSer	330
Db	1379	TCCCGCAGAGGGCGTCTGTTCGCATCCGCGCGGCGAACGAGGCCCGGAGTCGATC	1438
Qy	331	SerValAlaAsnThrAlaProTrrpValSerThrIleGlyAlaGlyThrLeuAspArg	350
Db	1439	GGTTCCCGCGCAGCGCGACCGCCCTCACCGTCGGCGCC	1480
Qy	351	PheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrrGlyGluSerLeuTyrrPro	370
Db	1480	-----	1480
Qy	371	GlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrrValThrGlyGlyAsp	390
Db	1481	-----GTCGACGACCAAGGAC	1495
Qy	391	LysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMet	410
Db	1496	AAG-----	1498
Qy	411	ValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAla	430
Db	1498	-----	1498
Qy	431	GlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAsp	450
Db	1498	-----	1498
Qy	451	ValHisLeuLeuProAlaThrLeuIleGlyTyrrThrGluSerValLeuLeuLysAlaTyrr	470
Db	1498	-----	1498
Qy	471	ValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArg	490
Db	1498	-----	1498
Qy	491	SerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSer	510
Db	1499	-----CTCGCGCACTTCTCTCCACCGCGCCCGCCCTCGCGCAGCGCGCC	1542
Qy	511	IleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrrpGln---	529
Db	1544	ATC---AAGCGGACGTCACCGCTCCCGGGTGGACATCACGGCGGCTCGCGGAGGGC	1600
Qy	530	-----AsnLeuGlyProThrGlyLeuProTyrrAspSerArgArg	542
Db	1601	AACGACATCGCCAGGAGGTCCGTCAGGACGCGCGCGC-----	1633
Qy	543	ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr	562
Db	1640	-----TACATGACCATCTCCGGCAGCTCGATGGCGACCCCGCGACGTCGCGGCGCGCG	1699
Qy	563	AlaLeuIleArgSerAlaTyrrProAsnTrrpSerProAlaIleLysSerAlaLeuMet	582
Db	1694	GCCCTCTGAAGCAGCAGCACCACGACTGACCTCCGCCGACTGAGGGCGCGCTC---	1755
Qy	583	ThrThrAlaAspLeuTyrrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla	602
Db	1751	-----ACCGGCTCCACCAAGGGC	1765
Qy	603	Gly-----ValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn	619

Db 1769 GGCAGGTACACCCCGTTCCGACGAGGTTCCGGCCGATCCAGCCGACAAAGCGCTCCAG 1828
Qy 620 ProGlyLeuVal-----TyrAsnIleGln----- 627
Db 1829 CAGACCGTGATCCGCGACCGCGTCTCGGTAGCTTCGGGTCCAGCAGTGGCGGCACACC 1888
Qy 628 -----ProVal---AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSer 643
Db 1889 GACGACGAGCGGTCCACCAAGCAGCTCACCACCGC---AACCTCGGC---ACCCAGGAC 1942
Qy 644 AspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnPro 663
Db 1943 GTCAGCTGAAGCTGACGTCCAGCGCCACCGACCCCAAGCGCAAGCGCGCCCGCGGCGC 2002
Qy 664 GlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGlu 683
Db 2003 TTCTTCACGTGGCGCCACACCGGTGACCGTCCCGCGCGGCGGCGACGCGCTCCGTCGAC 2062
Qy 684 MetIleThrArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnVal 703
Db 2063 ATG---ACCGCGACACCCCGCTCGCGGCGACCGTGGACGCGCGTACTCGGCGTACGTG 2119
Qy 704 LysAlaProGluGly 708
Db 2120 GTCGCCACGCGCGC 2134
RESULT 11
US-09-000-016-1
; Sequence 1, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000.016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914

ORIGINAL SOURCE: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores:
Pred. No.: 1,48e-19 Length: 2809
Score: 297.00 Matches: 155
Percent Similarity: 31.03% Conservative: 70
Best Local Similarity: 21.38% Mismatch: 224
Query Match: 7.39% Indels: 276
DB: Gaps: 22
US-09-806-767-2 (1-775) x US-09-000-016-1 (1-2809)
Qy 36 SerGluThrAlaLysThrPheAlaSer---LysPheAspTrpHisLeuSerPheLeuGln 54
Db 632 GCGGACGCGCGCGCTGGTCCGACGCGGCAAGCTCGACCGCGGCTCTTCGACATCACC 691
Qy 55 GluAlaValLeuGlyValGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 74
Db 692 GAACCTCGGCAAGCGCGGACCCCGCACTCCCGAAGAACAGGACTGAAGGTCATCTCGCGC 751
Qy 75 TyrGlySerAlaIleGluGlyPheAlaAlaGln----- 85
Db 752 TACAGGCGCGCGACG 811
Qy 86 ---LeuThrGluSerGluAlaGluLeuArgTyrSerProGluValAlaValArg 104
Db 812 ACCCTGACGTCCCTGAACGCGGACGCGGTGGCGGACCGCGACGAGGACGCGTGGAGCTG 871
Qy 105 ProAspHisValLeuGlnValGlnThrThrTyrSer---TyrLysPheLeuGlyLeuAsp 123
Db 872 TGGGACGCGCGTCAACACGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 931
Qy 124 Gly-----PheGlyAsnSerGlyValTyr 131
Db 932 GGGGTCCGACG 991
Qy 132 SerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyValTyrPro 151
Db 992 TCCGCGCGGTACGACGCGGCAAGGCGGTGAAGATCGCGGTCTGGACACCGGTGTC----- 1045
Qy 152 GluSerProSerPheAspThrGlyMetProSerIleProArgLysTyrLysGlyIle 171
Db 1046 -----GACACGAGCATCCGACCTG-----AAGGCGCGG 1075
Qy 172 CysGlnGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIleGlyAlaArg 191
Db 1076 GTGACCGCGTCCAGAGACTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1108
Qy 192 PhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArg 211
Db 1108 ----- 1108
Qy 212 GluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGly 231
Db 1109 -----GGCGCGCGGACAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1159
Qy 232 GlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMet 251
Db 1160 GGCACGCGGCG-----GCCAGTCCAAAGGCGAGTACAGGCGCGCTC 1198
Qy 252 AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys---TyrSer 270

Db 1199 GCACCCGGCGCGGATCTCAACGGCAAGTCTCTGACGACTCCGGTTTCGGCGCAGCAG 1258
Qy 271 SerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeu 290
Db 1259 TCCGGCATCTCCCGCATGAGTGGCGCGCGCGCAGGCGCCGCGTCTCAACATG 1318
Qy 291 SerLeuGlyGlyPheProIleProLeuTyAspAspThrIleAlaIleGlyThrPheArg 310
Db 1319 AGCTGGCGCGCATGACACACCGGACCGACCGCTGGAGCGCGGTTCGACAAAGCTG 1378
Qy 311 AlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSer 330
Db 1379 TCCGCCGAGAGGCGTCTTCGCGCATCGCGCGCGCAACGAGGCGCGGAGTCTGATC 1438
Qy 331 SerValAlaAsnThrAlaProTTPValSerThrIleGlyValaclyThrLeuAspArg 350
Db 1439 GGTTCGCGCGCGCAGCGCGCGCGCTCACGTCGGCGCC----- 1480
Qy 351 PheProAlaValArgLeuAlaAsnGlyLysLeuLeuTyrglyGluSerLeuTyPro 370
Db 1480 ----- 1480
Qy 371 GlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyValThrGlyGlyAsp 390
Db 1481 -----GTCCGACGACAAAGGAC 1495
Qy 391 LysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMet 410
Db 1496 AAG----- 1498
Qy 411 ValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAla 430
Db 1498 ----- 1498
Qy 431 GlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAsp 450
Db 1498 ----- 1498
Qy 451 ValHisLeuLeuProAlaThrLeuIleGlyTyThrGluSerValLeuLeuLysAlaTy 470
Db 1498 ----- 1498
Qy 471 ValAsnAlaThrValLysProLysAlaArgIlePheGlyGlyThrValIleGlyArg 490
Db 1498 ----- 1498
Qy 491 SerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSer 510
Db 1499 -----CTCGCGGACTTCTCTCCACCGCGCGCGCTCGCGCAGCGCGCC 1543
Qy 511 IleLeuLysProAspMetIleAlaProGlyValAsnIleAlaAlaIleProGln----- 529
Db 1544 ATC---AAGCGGAGCGTCACCGTCTCGCGCGTGCATCATCGCGCGCTCGCGGAGGCG 1600
Qy 530 -----AsnLeuGlyProThrGlyLeuProTyraSpSerArgArg 542
Db 1601 AACGACATCGCGCAGGAGTCTGAGGAGCGCGCGCG----- 1639
Qy 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1640 -----TACATGACCATCTCGGACGTCGATGGGACCGCGCGCGCGCGCGCGCG 1693
Qy 563 AlaLeuIleArgSerAlaTyProAsnTTPSerProAlaAlaIleLysSerAlaLeuMet 582
Db 1694 GCCTCTCTGAAGCAGCAGCAGCCCGACTCGACCTCGCGCACTGAAGGCGCGCTC--- 1750
Qy 583 ThrThrAlaAspLeuTyAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db 1751 -----ACCGGCTCCACCAAGGCG 1768
Qy 603 Gly-----ValPheAlaIleGlyValaclyHisValAsnProGlnLysAlaIleAsn 619
Db 1769 GGCAAGTACACCGGTCGAGCAGGGTTCGGGCGGATCCAGCGCGCAGCAAGCGCTCCAG 1828

Qy 620 ProGlyLeuVal-----TyrAsnIleGln----- 627
Db 1829 CAGACCGTATCGCGCAGCCCGGTCTCGGTAGCTTCGGCTGCCAGTGGCGCCACACC 1888
Qy 628 -----ProVal---AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSer 643
Db 1889 GACGACGAGCGGTTCACCAAGCAGCTACCTACCGC---AACCTCGGC---ACCCAGGAC 1942
Qy 644 AspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnPro 663
Db 1943 GTCAGCTGAGCTGACGTGCGACCGCCACCGACCCCAAGGCAAGGCGCGCGCGCGGC 2002
Qy 664 GlyPheSerLeuAsnTyProSerIleAlaValIlePheLysArgGlyLysThrGlu 683
Db 2003 TTCCTTCAGCTGGCGCCACCGACCGTACCGTCCCGCGCGCGCGCAGCGCTCGCTCGAC 2062
Qy 684 MetIleThrArgValThrAsnValGlySerProAsnSerIleTySerValAsnVal 703
Db 2063 ATG---ACCGCGCAGACCGCGCTCGCGCGCAGCGTGCAGCGCGCTACTCGGCGTACGTG 2119
Qy 704 LysAlaProGlyGly 708
Db 2120 GTCGCCACCGCGCGC 2134
RESULT 12
US-09-514-340-1
Sequence 1, Application US/09514340
Patent No. 6361987
GENERAL INFORMATION:
APPLICANT: AKIRA ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE D
ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN: <Unknown>
FEATURE:


```

; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1

Alignment Scores:
Pred. No.: 1,48e-19 Length: 2809
Score: 297.00 Matches: 155
Percent Similarity: 31.03% Conservative: 70
Best Local Similarity: 21.38% Mismatches: 224
Query Match: 7.39% Indels: 276
DB: 4 Gaps: 22

US-09-806-767-2 (1-775) x US-09-514-340-1 (1-2809)
Qy 36 SerGluThrAlaLysThrPheAlaSer---LysPheAspTrpHisLeuSerPheLeuGln 54
Db 632 GCGAGCGCCCGCGGTGTGTCGCGAGCGGCAAGCTCGACCGCGGCTCTTCGACATCACC 691
Qy 55 GluAlaValLeuGlyValGluGluGluGluGluGluProSerSerArgLeuLeuTyrSer 74
Db 692 GAATCGGCAAGCGCGGACCCCACTCCAGAAACAGGACTGAGGTCTATCGTCGCG 751
Qy 75 TyrGlySerAlaIleGluGlyPheAlaAlaGln----- 85
Db 752 TACCAGGCGCGCACGCGCGCGCAAGCGCGAGGTCCGCGAAGCGGCGAACTCCGCGCG 811
Qy 86 ---LeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArg 104
Db 812 ACCCTGACGTCCCTGAACGCGGAGCGGTGCGGACCCCGCACGCGCGTCCGAGCTG 871
Qy 105 ProAspHisValLeuGlnValGlnThrTyrSer---TyrTyrPheLeuGlyLeuAsp 123
Db 872 TGGGACCCCTGACCAACGCGGAGCGGACCGCGCTCGGATCGCCACGCTCTGGCTGGAC 931
Qy 124 Gly-----PheGlyAsnSerGlyValTrp 131
Db 932 GGGGTCCGAGGCGCGCTCGACACGTCGTCGGGACAGATCGCGCCCAAGGGGTGG 991
Qy 132 SerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyValTrpPro 151
Db 992 TCGCGCGCTACACGCGCAAGGCGGTGAAGATCGCGTCTCGACACCGGTGTC----- 1045
Qy 152 GluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLysGlyIle 171
Db 1046 -----GACACGAGCCATCCGAGACTG-----AAGGGCCGG 1075
Qy 172 CysGlnGluGlyGluSerPheSerSerSerCysAsnArgLysLeuIleGlyAlaArg 191
Db 1076 GTGACCGCGTCCAGAACTTCAACGCGCGCGCC----- 1108
Qy 192 PhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArg 211
Db 1108 ----- 1108
Qy 212 GluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGly 231
Db 1109 -----GGCGCGCGCGCAAGGTGGCGCACCGCACCCACGCTCGCTCGATCGCGGG 1159
Qy 232 GlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMet 251
Db 1160 GGCACGCGG-----GCCAGTCCAGGGCGAAGTACAAAGGGCGTC 1198
Qy 252 AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCys---TyrSer 270
Db 1199 GCACCGCGCGCGGATCCCTCAACGCGCAAGGTCTCTGACGACTCCGCTTCGCGCGAC 1258
Qy 271 SerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeu 290

```

```

Db 1259 TCGGCATCTCTCGCGCATGGAGTGGCGGCGCGCAGCGGCCGCGTGTCAACATG 1318
Qy 291 SerLeuGlyGlyPheProIleProLeuTyrAspThrIleAlaIleGlyThrPheArg 310
Db 1319 AGCTGGCGCGCATGGACACACCGGAGACCGCCCTGAGGCGGGGTGCAAGAGCTG 1378
Qy 311 AlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSer 330
Db 1379 TCGCGCGGAGAGGGGCTCTGTCGCTCGCCATCGCGCGCGCAACGAGGCGCGGAGTCGATC 1438
Qy 331 SerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArg 350
Db 1439 GGTTCGCGCGCGCAGCGCGGACCGCCCTCACGCTCGCGCC----- 1480
Qy 351 PheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrPro 370
Db 1480 ----- 1480
Qy 371 GlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyAsp 390
Db 1481 -----GTGACGCAAGGAC 1495
Qy 391 LysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMet 410
Db 1496 AAG----- 1498
Qy 411 ValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAla 430
Db 1498 ----- 1498
Qy 431 GlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGluAspSerIleAsp 450
Db 1498 ----- 1498
Qy 451 ValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyr 470
Db 1498 ----- 1498
Qy 471 ValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArg 490
Db 1498 ----- 1498
Qy 491 SerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSer 510
Db 1499 -----CTGCGCGACTTCTCTCCACCGCGCCCGCTCGGCGAGCGCGCC 1543
Qy 511 IleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGln--- 529
Db 1544 ATC---AAGCCGAGCTCACCGCTCCGCGGTGGACATCACGCGCGCTCGGCGAGGGC 1600
Qy 530 -----AsnLeuGlyProThrGlyLeuProTyrAspSerArgArg 542
Db 1601 AACGACATCGCGCAGGAGTGGTGAGGACCGCGCGC----- 1639
Qy 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyLeuThr 562
Db 1640 -----TACATGACCATCTCCGCGACGTCGATGCGACCGCCGCGCGCGCGCGCG 1693
Qy 563 AlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMet 582
Db 1694 GCCCTCTTGAAGACGACGACACCCGACTGGACCTCCGCGCGACTGAAGGGCGCGCTC--- 1750
Qy 583 ThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db 1751 -----ACCGGCTCCCAAGGGC 1768
Qy 603 Gly-----ValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn 619
Db 1769 GGCAGAGTACACCCCGTTCGACAGGGTTCGCGCGGATCCAGGCGCGCAAGGCGCTCCAG 1828
Qy 620 ProGlyLeuVal-----TyrAsnIleGln----- 627

```

Db 1829 CAGACGGTATCGCCGACCCCGGTCTCGGTGAGCTTCGGCTCCAGCAGTGGCCGACACC 1888
 QY 628 -----ProVal---AspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSer 643
 Db 1889 GACGACGAGCGGTCAACAGCAGCTGACCTACCGC---AACCTCGGC---ACCAGGAC 1942
 QY 644 AspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArgLysAsnPro 663
 Db 1943 GTCACGCTGAAGCTGACGTCGACCGCCACCGACCCCAAGGCGGCGGCCGCGGC 2002
 QY 664 GlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGlu 683
 Db 2003 TTCCTTCAGCTGGCGCCACACGCTGACCGTCCCGCGGCGGCGGCGGCTCCGTCGAC 2062
 QY 684 MetIleThrArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnVal 703
 Db 2063 ATG---ACGCGGACACCGGCTCGCGCGCACGGTGGACGGCGGTACTCGCGGTACGTG 2119
 QY 704 LysAlaProGluGly 708
 Db 2120 GTCGCGACGGCGGC 2134

RESULT 13

US-09-255-502-1
 ; Sequence 1, Application US/09255502
 ; Patent No. 6218165
 ; GENERAL INFORMATION:
 ; APPLICANT: Estell, David
 ; TITLE OF INVENTION: Mutant Proteins Having Lower Allergenic Responses in
 ; TITLE OF INVENTION: Humans and Methods for Constructing, Identifying and
 ; TITLE OF INVENTION: Producing Such Proteins
 ; FILE REFERENCE: GC 527-D2
 ; CURRENT APPLICATION NUMBER: US/09/255,502
 ; CURRENT FILING DATE: 1998-02-23
 ; PRIOR APPLICATION NUMBER: 09/060,872
 ; PRIOR FILING DATE: 1998-04-15
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1494
 ; TYPE: DNA
 ; ORGANISM: Bacillus amyloliquefaciens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (36)..(1244)
 ; NAME/KEY: mat_peptide
 ; LOCATION: (417)..(1241)
 US-09-255-502-1

Alignment Scores:

Pred. No.: 1.89e-17 Length: 1494
 Score: 271.50 Matches: 126
 Percent Similarity: 30.73% Conservative: 63
 Best Local Similarity: 20.49% Mismatches: 177
 Query Match: 6.76% Indels: 249
 DB: 3 Gaps: 17

US-09-806-767-2 (1-775) x US-09-255-502-1 (1-1494)

QY 6 PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerGluIleLeuGlnLys 25
 Db 129 TTTGCTTTAGCGTTAATCTTTACGATGGCGGTTCGCGACACATCTCT----- 176
 QY 26 GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLys 45
 Db 177 -----GCCGCGCGCAGGGAATCAACGGGGAAAGAAATATATTGTCGGG 224
 QY 46 PheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGlu 65
 Db 225 TTTAAACAGACAAATGACGACGATGAGCGCGCTAAGAGAGAGATGTCATTCTGAAAAA 284
 QY 66 GluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaGln 85

Db 285 GCGCGGAAGTCAAAAGCAATTCAAAATAT-----GTAGACGACGCTTCAGCTACA 335
 QY 86 LeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArgPro 105
 Db 336 TTTAAACGAAAAGCTCTAAAGAATTGAAAAAGACCCGAGCGTCCGTTACGTTGAAGAA 395
 QY 106 AspHisValLeuGlnValGlnThrThrTyrSerTyrLysPhe---LeuGlyLeuAspGly 124
 Db 396 GATCAGCTAGCACAT-----GCGTACGGCGAGTCCGTCCTTACGGCTATCACAA 446
 QY 125 PheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyVal 144
 Db 447 ATTAAAGCCCTGCTCTGCACCTCAAGGCTACACTGGATCAAAATGTTAAAGTAGCGGT 506
 QY 145 LeuAspThrGlyValTrpProGluSerProSerPheAspThrGlyMetProSerIle 164
 Db 507 ATCGACGCGGTATCATCTCTCTCATCTCTGATTTAAAGGTAGCAGGCGGAGCCGACG 566
 QY 165 ProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsn 184
 Db 566 ----- 566
 QY 185 ArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGlu 204
 Db 566 ----- 566
 QY 205 GluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThr 224
 Db 567 -----GTTCCCTTCTGAAACAAATCCTTTCCAGACCAACACTCTCACCGAACT 614
 QY 225 HisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGly 244
 Db 615 CAGGTGCGCGCACAGTTCCG-----GCTCTTAATAACTCA 650
 QY 245 AlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrp 264
 Db 651 ATCGGTGTATTA---GGCGTTCGCCAAGCGCATCACTTTCAGCTGTAAGTTCCTCGT 707
 QY 265 PheAsnGlyCys---TyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAsp 283
 Db 708 GCTGACGGTTCGCGCCAATACAGCTGGATCAATACGGAATCGATCGCGCGCATCAAC 767
 QY 284 LysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThr 303
 Db 768 AATATGGAGCTTATTAAACATGAGCTTCGCGGA-----CCTTCGGTTCCTCGCT 818
 QY 304 IleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGly 323
 Db 819 TTTAAAGCGCGCAGTTGATAAAGCCGTCGATCCGCGTCCGATGTCGTCGCGCAGCCGT 878
 QY 324 AsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGly 343
 Db 879 AACGAAGGCACCTCCGCGCAGCTCA-----AGCACAGTGGGC 914
 QY 344 AlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeu 363
 Db 914 ----- 914
 QY 364 TyrGlyLysSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluVal 383
 Db 915 -----TACCTCGTAAA----- 926
 QY 384 IleTyrValThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArg 403
 Db 926 ----- 926
 QY 404 GluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLys 423
 Db 926 ----- 926
 QY 424 GlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsn 443

Db 926 ----- 926
Qy 444 GlnGluAspSerIleAspValHisLeuLeuProAlaThrLeu---IleGlyTyrThr 462
Db 927 -----TACCCTCTGTCTATTTCAGTAGCGGTGT 956
Qy 463 GluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIle 482
Db 957 GACAGC----- 962
Qy 483 PheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArg 502
Db 963 -----AGCAACCAAGAGCATCTTTCTCAAGCGTA 992
Qy 503 GlyProSerIleuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsn 522
Db 993 GGACCTGAGCTT-----GATGTCATGGCAGCTGGCGTATCT 1028
Qy 523 IleIleAlaIatProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArg 542
Db 1029 ATCAAGACGCTTCTCTGGAAC----- 1052
Qy 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1053 ---AAATACGGCGGTACAAACGGTACGTCAATGGCATCTCCGACGTTGCCGAGCGGT 1109
Qy 563 AlaLeuIleArgSerAlaTyrProAsnTyrSerProAlaIleLysSerAlaLeuMet 582
Db 1110 GCCTTGATCTTCTTAAGCACCCGAACTGGACAAACACTCAAGTCCGCGACGATTTAGAA 1169
Qy 583 ThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db 1170 AACACCACT-----ACAAACTTGGT 1190
Qy 603 GlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAla 617
Db 1191 GATTCTTCTACTGGAAGGCTGATCAAGTACAGCGCA 1235

RESULT 14
5472855-1
; Patent No. 5472855
; APPLICANT: CARTER, PAUL J.; WELLS, JAMES A.
; TITLE OF INVENTION: SUBSTRATE ASSISTED CATALYSIS
; NUMBER OF SEQUENCES: 31
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/287,964
; FILING DATE: 22-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 90,902
; FILING DATE: 12-JUL-1993
; APPLICATION NUMBER: 823,039
; FILING DATE: 14-JAN-1992
; APPLICATION NUMBER: 35,652
; FILING DATE: 06-APR-1987
; APPLICATION NUMBER: 334,081
; FILING DATE: 04-APR-1989
; APPLICATION NUMBER: 127,134
; FILING DATE: 01-DEC-1987
; APPLICATION NUMBER: 846,627
; FILING DATE: 01-APR-1986
; APPLICATION NUMBER: 858,594
; FILING DATE: 30-APR-1986
; APPLICATION NUMBER: 614,612
; FILING DATE: 29-MAY-1984
; APPLICATION NUMBER: 614,615
; FILING DATE: 29-MAY-1984
; APPLICATION NUMBER: 614,617
; FILING DATE: 29-MAY-1984
; APPLICATION NUMBER: 614,491
; FILING DATE: 29-MAY-1984
; SEQ ID NO.1:
; LENGTH: 1497
5472855-1

Alignment Scores: 1.88e-17 Length: 1497
Pred. No.: 271.50 Matches: 126
Score: 20.73% Conservativeness: 63
Percent Similarity: 20.49% Mismatches: 177
Best Local Similarity: 6.76% Indels: 249
Query Match: 6 Gaps: 17
DB: 6
US-09-806-767-2 (1-775) x 5472855-1 (1-1497)
Qy 6 PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLys 25
Db 129 TTTGCTTTAGCGTTAAATCTTTTACGATGGCGTTCGGCAGCACATCTCTCT----- 176
Qy 26 GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLys 45
Db 177 -----GCCAGCGGCGCAGGGAATCAACCGGGAAGAAATATATTTGTCGGG 224
Qy 46 PheAspTyrHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGlu 65
Db 225 TTTAAACAGACAATGAGCAGCATGAGCGCGCTAAGAGAAAGATGTCATTTCTGAAAAA 284
Qy 66 GluProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGln 85
Db 285 GCGCGGAAAGTGCAAAAGCAATTCAAATAT-----GTAGACGCGAGCTTCAGCTACA 335
Qy 86 LeuThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArgPro 105
Db 336 TTAACGAAAAAGCTGTAAAGAAATGAAAAAGACCCGAGCGTTCGTTAGTTGAAGAA 395
Qy 106 AspHisValLeuGlnValGlnThrTyrSerTyrLysPhe---LeuGlyLeuAspGly 124
Db 336 GATCAGGTAGCACAT-----GCGTACGCGCAGTCCGTCCTTACGGGATACAA 446
Qy 125 PheGlyAsnSerGlyValTyrSerLysSerArgPheGlyGlnGlyThrIleIleGlyVal 144
Db 447 ATTAAGAGCCCTGCTCTGCACTCTCAAGGCTACACTGGATCAATGTTAAAGTAGCGT 506
Qy 145 LeuAspThrGlyValTyrProGluSerProSerPheAspThrGlyMetProSerIle 164
Db 507 ATCGACAGCGGTATCGATTCTTCTCATCTGATTAAAGGTAGCAGCGGAGCCAGCATG 566
Qy 165 ProArgLysTyrLysGlyLeuCysGlnGluGlyLeuSerPheSerSerSerCysAsn 184
Db 566 ----- 566
Qy 185 ArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGlu 204
Db 566 ----- 566
Qy 205 GluSerProAsnMetProArgGluTyrIleSerAlaArgSerThrGlyHisGlyThr 224
Db 567 -----GTTCCCTTCTGAAACAAATCCTTTCCAAGACAACTCTCAGCGAACT 614
Qy 225 HisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGly 244
Db 615 CACGTTGCCGCGCACAGTTGCG-----GCTCTTAAATACTCA 650
Qy 245 AlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrp 264
Db 651 ATCGGTGTATTA---GGCGTTGCCGCAAGCGATCACTTTACGCTGTAAAAAGTTCTCGGT 707
Qy 265 PheAsnGlyCys---TyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAsp 283
Db 708 GCTGACGTTCCGGCCCAATACAGCTGGATCATTAACGGAATCGAGTGGCGATCGCAAC 767
Qy 284 LysValAspValLeuSerLeuSerIleGlyPheProIleProLeuTyrAspAspThr 303
Db 768 AATATGACGATTATTAAATGACATGAGCTCGCGGA-----CCTTCTGGTCTGCTGCT 818
Qy 304 IleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGly 323
Db ----- 323

```

Db      819  TTAAGAGCGCAGTTGATAAGACCGTGTGTCATCGCGCTCGTAGTCTGTCGCGCAGCCGGT 878
Qy      324  AsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGly 343
Db      879  AACGAGGCACTTCGCGCAGCTCA-----AGCACAGTGGGC 914
Qy      344  AlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeu 363
Db      914  ----- 914
Qy      364  TyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluVal 383
Db      915  -----TACCTGTGTA----- 926
Qy      384  IleTyrValThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArg 403
Db      926  ----- 926
Qy      404  GluGluIleArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSerGluLys 423
Db      926  ----- 926
Qy      424  GlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluLeuAsn 443
Db      926  ----- 926
Qy      444  GlnGluGluAspSerIleAspValHisLeuLeuProAlaThrLeu-----IleGlyTyrThr 462
Db      927  -----TACCCTTCGTCTATGTCAGTAGCGCTGTT 956
Qy      463  GluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIle 482
Db      957  GACAGC----- 962
Qy      483  PheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArg 502
Db      963  -----AGCAACCAAGAGCATCTTTCTCAAGCGTA 992
Qy      503  GlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsn 522
Db      993  GGACCTGAGCTT-----GATGTCATGGCACCTGGCGTATCT 1028
Qy      523  IleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgArg 542
Db      1029  ATCAAGACGAGCTTCTCGAATC----- 1052
Qy      543  ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db      1053  ---AAATACGGGCGGTACAAACGGTACGTCAATGGCATCTCCGACCGTTCGCGAGCGGT 1109
Qy      563  AlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMet 582
Db      1110  GCATTGATTTCTTAAGCACCCGGAACCTGACAAACACTCAAGTCGCGACGAGTTAGAA 1169
Qy      583  ThrThAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db      1170  AACACCACT-----ACAAACTGGT 1190
Qy      603  GlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAla 617
Db      1191  GATTCITCTACTATGAAAAAGGCTGATCAACAGTACAGCGGCA 1235

RESULT 15
US-08-069-863-1
; Sequence 1, Application US/08069863
; Patent No. 5470733
; GENERAL INFORMATION:
; APPLICANT: BRYAN, Philip N
; APPLICANT: ALEXANDER, Patrick
; APPLICANT: STRAUSSBERG, Susan L
; TITLE OF INVENTION: CALCIUM FREE SUBSTITILISIN MUTANTS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/069,863
; FILING DATE: 01-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 028755-016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 450..1599
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 772..1599
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 450
; OTHER INFORMATION: /note="Amino Acid Val at position
; OTHER INFORMATION: 450 is fMet."
; US-08-069-863-1

Alignment Scores:
Pred. No.: 2,76e-17 Length: 1868
Score: 271.50 Matches: 126
Percent Similarity: 30.73% Conservative: 63
Best Local Similarity: 20.49% Mismatches: 177
Query Match: 6.76% Indels: 249
DB: 1 Gaps: 17

US-09-806-767-2 (1-775) x US-08-069-863-1 (1-1868)
Qy      6  PhePheLeuCysIleIlePheLeuLeuPheCysSerSerSerGluIleLeuGlnLys 25
Db      483  TTTCCTTTAGCGTTAATCTTTACGATCGCGTTCGGCAGCACATCCTCT----- 530
Qy      26  GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLys 45
Db      531  -----GCCACGGCGGCGAGGAATCAACCGGGGAAAGAAATATATTTGTCGGG 578
Qy      46  PheAspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGlu 65
Db      579  TTTAAACAGACATGACGACGATGACGCGCTAAGAGAAAGATGTCATTTCGAAAAA 638
Qy      66  GluProSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGln 85
Db      639  GGCGGGAAAGTGCAAAAGCAATTCAAATAT-----GTAGACGCGAGCTTCAGCTACA 689
Qy      86  LeuThrGluSerGluAlaGluIleLeuArgTyrSerProGlnValValAlaValArgPro 105
Db      690  TTAACGAAAAAGCTGTAAAGAAATTTGAAAAAGACCCGAGCGCTTACGTTGAAGAA 749
Qy      106  AspHisValLeuGlnValGlnThrThrTyrSerTyrLysPhe---LeuGlyLeuAspGly 124

```

```
Db 750 GATCAGGTAGCACAT-----GCGTACGGCGAGTCGCTGCTTACGGCGTATCACAA 900
Qy 125 PheGlyAsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyVal 144
Db 801 ATTAAGCCCTCTCTGCACTCTCAAGGCTACACTGGATCAAAATGTTAAAGTACGGTT 860
Qy 145 LeuAspThrGlyValTrpProGluSerProSerPheAspAspThrGlyMetProSerIle 164
Db 861 ATCGACAGCGGTATCGATTCTCTCATCTCTGATTAAAGGTAGCAGCGGAGCCAGCATG 920
Qy 165 ProArgLysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerCysAsn 184
Db 920 ----- 920
Qy 185 ArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGlu 204
Db 920 ----- 920
Qy 205 GluSerProAsnMetProArgGluTyriIleSerAlaArgAspSerThrGlyHisGlyThr 224
Db 921 -----GTTCTCTCTGAAACAATAATCTCTTCCAGACACAACACTCTCAGCGAACT 968
Qy 225 HisThrAlaSerThrValGlyLysSerSerValSerMetAlaAsnValLeuGlyAsnGly 244
Db 969 CACGTTGCCGGCACAGTTGG-----GCTCTTAATAACTCA 1004
Qy 245 AlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyriLysValCysTrp 264
Db 1005 ATCGGTGTATTA---GGCGTTGGCCAGCGCACTCACTTACGCTGTATAAAGTTCTCGGT 1061
Qy 265 PheAsnGlyCys---TyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAsp 283
Db 1062 GCTGACGTTCCGGCCAAATACACTGATCATTAACGGAATCGAGTGGCGGATCGCAAC 1121
Qy 284 LysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyriAspAspThr 303
Db 1122 AATATGACGCTTATAACATGACCTCGCGGA-----CCTTCTGTTCTGCTGCT 1172
Qy 304 IleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGly 323
Db 1173 TTAAGAGCGGCGAGTGTATAAGCGGTTCATCGCGGTGCTGTAGTCTGCGGCGCGCGGT 1232
Qy 324 AsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGly 343
Db 1233 AACGAGGCGACTTCCGGCAGCTCA-----AGCACAGTGGGC 1268
Qy 344 AlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeu 363
Db 1268 ----- 1268
Qy 364 TyrGlyGluSerLeuTyriProGlyLysGlyIleLysAsnAlaGlyArgGluValGluVal 383
Db 1269 -----TACCCTGGTAAA----- 1280
Qy 384 IleTyriValThrGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArg 403
Db 1280 ----- 1280
Qy 404 GluGluIleArgGlyLysMetValIleCysAspArgGlyValAlaAsnGlyArgSerGluLys 423
Db 1280 ----- 1280
Qy 424 GlyGluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsn 443
Db 1280 ----- 1280
Qy 444 GlnGluAspSerIleAspValHisLeuLeuProAlaThrLeu---IleGlyTyriThr 462
Db 1281 -----TACCCTTCTGTCATTCAGTAGCGCGCTGT 1310
Qy 463 GluSerValLeuLeuLysAlaTyriValAlaAsnAlaThrValLysProLysAlaArgIleIle 482
```

```
Db 1311 GACAGC----- 1316
Qy 483 PheGlyGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAlaArg 502
Db 1317 -----AGCAACCAAGAGAGCATCTTTCTCAAGCGTA 1346
Qy 503 GlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsn 522
Db 1347 GGACCTGAGCTT-----GATGTCATGGCACCTGCGCTATCT 1382
Qy 523 IleIleAlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyriAspSerArgArg 542
Db 1383 ATCCAAAGCAGCTTCTCTGGAAC----- 1406
Qy 543 ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr 562
Db 1407 -----AAATACGGGGCGGTACAACGGTACGTCAATGGCATCTCCGACGTTGCGGAGCGGCT 1463
Qy 563 AlaLeuIleArgSerAlaTyriProAsnTrpSerProAlaAlaIleLysSerAlaLeuMet 582
Db 1464 GCITTTGATTCTTTCTAAGCACCGGAACTGGACAAACACTCAAGTCCGCGACGAGTTAGAA 1523
Qy 583 ThrThrAlaAspLeuTyriAspArgGlnGlyLysAlaIleLysAspGlyAsnLysProAla 602
Db 1524 AACACCACT-----ACAAAACTTGGT 1544
Qy 603 GlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAla 617
Db 1545 GATTCCTTTCTACTATGAAAAGGCGTCAATCAAGTACAGCGCGCA 1589
```

Search completed: February 3, 2004, 04:37:39
Job time : 147 secs

; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1065

Alignment Scores:

Pred. No.:	3,138-198	Length:	2255
Score:	1742.00	Matches:	370
Percent Similarity:	63.4%	Conservative:	115
Best Local Similarity:	48.43%	Mismatches:	241
Query Match:	43.35%	Indels:	38
DB:	10	Gaps:	15

US-09-806-767-2 (1-775) x US-09-938-842A-1065 (1-2265)

Qy	11	IlePheLeuLeuPheCysSerSerSerSerSerGlnThrThrThrIleVal	30
Db	52	CTCTCTCTCTCTCTCACACACACCC	96
Qy	31	GlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTprHisLeu	50
Db	97	CGCGTCAAT--CACTCGATAAACCGGAATCATCTCTCACTCCACACGATTGGTACACA	153
Qy	51	SerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSerSerArg	70
Db	154	TCTCACTCAATCA	177
Qy	71	LeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaGlnLeuThrGluSerGlu	90
Db	178	CTTCTCTACACTACACACCTCTCCATCGCTTCCGGTACCTCGACTCCACCGAA	237
Qy	91	AlaGlu--IleLeuArgTyrSerProGluValValAlaValArgProAspHisValLeu	109
Db	238	GCGGATCTCTCTCTCCCTCAACCAATCTCGATATCTTCGSAAGATCTCTCTATC	297
Qy	110	GlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGly--PheGlyAsnSer	128
Db	298	ACATTCACACTACGCGTACTCTCGATTTCTCGTCTCAATCCGATTCGGTGTTCCAC	357
Qy	129	GlyValTprSerLysSerArgPheGlyGlnGlyThrIleLeuValLeuAspThrGly	148
Db	358	GATCTCGGTTCTCTCT-----AACGCGTTATCATCGGAGTTTAGATACTGGC	408
Qy	149	ValTprProGluSerProSerPheAspThrGlyMetProSerIleProArgLysTpr	168
Db	409	GTTTGGCCTGAATCTAGAAAGCTTCGATGATCTGATATCGCTGAGATTCTCTTAATGG	468
Qy	169	LysGlyLysCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLysLeuLe	188
Db	469	AAAGGAGATGTGAATCTCGTTCCGATTCCGAATCCAAAGTTGTGTAAACAGACGTTATC	528
Qy	189	GlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsn	208
Db	529	GGAGCTAGAGCTTCCAAAGAGATTCAAATGGCTTCTGGTGGTGGTTTTTCAGTTAG	588
Qy	209	MetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSer	228
Db	589	-----CGTGAATCTGTTCTCTCGTGTGTTGACGACATCGAACACATCTCAACT	642
Qy	229	ThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAla	248
Db	643	ACCGCGCGGGATCCCGCTAGAACCGTAGCTTCTCGGTATCGCGCGGTCACGGCC	702
Qy	249	ArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTprPheAsnGlyCys	268
Db	703	AGAGGTATGCCACTCGTCTCGTGTGCTACTTATAAGTTTGTGAGTACTGGTTGT	762
Qy	269	TyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeu	288
Db	763	TTTGGATCTGATATACTAGTCTATGGATCGAGCTATACTTGTATGGTGTGTATGTCCT	822
Qy	289	SerLeuSerLeuGlyGlyPheProIleProLeuTyrAspThrIleAlaIleGlyThr	308
Db	823	TCGTTATCTCTGGTGGTGGTCTGCTCGTATTATCGGATACGATGTGATTGGAGCG	882

309	PhaArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle	328
883	TTTTTCGGCTATGAGAGAGGTGTTTTTGTCCTTTGCTCGTGGTAATAGTGGCTCTACT	942
329	GluSerSerValAlaAsnThrAlaProTtpValSerThrIleGlyValaGlyThrLeuAsp	348
943	AGAGCTCTCTGTTGCCAATGTTGCTCTTGGTTATGACTGTGTGGTCTGCTACTTTAGAT	1002
349	ArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeu	368
1003	AGAGATTTTCGGCTTTTTCGAATCTCGGTAAACGGGAACGACTTACCGGTGTTCCGCTG	1062
369	TyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGly	388
1063	TATAGCGGTGTCAGCAATG-----GGACCAAGACCGCTTGAATGGTTTATATAAAGGG	1116
389	GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGly	408
1117	AATAGTAGTTCGAGTAATCTTTGTTTACCTGGTTCGTTGATTCGAGTATGTTGCTGGG	1176
409	LysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLys	428
1177	AAGATTGTTGTTGATAGAGGTGTTAATGCTAGAGTTGAGAAAGAGCGTGTGGTTAGA	1236
429	GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSer	448
1237	GATGCTGCTGTTTAGGATGATAAATGGCGAAATACCTCGCGAGTGCAGAGGAGCTTGTG	1296
449	IleAspValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLeuLys	468
1297	CGCGATAGTCATTGCTTCCCGGATCGCTGTAGGGAAGAGACTGGTGATTTACTTTAGG	1356
469	AlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGlyThrValIle	488
1357	GAGTAAGTTAAGTCAGATTCTAAACCAACCGCTCTCTTCTGTTTAAAGGAACGGTTCCT	1416
489	GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsn	508
1417	GACGTTAAGCGCTCTCTGTTGTTGCTGCTTTTAGCTCGAGAGCTCGTAATCTGTTACT	1476
509	ProSerIleLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTppPro	528
1477	CCTGAATCTGAAGCCTGATGTTATTTGGTCCCTGGAGTTAATATTTTGGCTGGTGGCT	1536
529	GlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgArgValAsnPheThrValMet	548
1537	GACGCTATTGGTCTCTACTGCTCTTGACAGCATCTAGGAGGACTCAGTTCAACATCATG	1596
549	SerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAla	568
1597	TCAGGTACGTCATGTGTCATGCCACACATCAGTGGTATTAGCGGTCTTTTGAAGACAGCT	1656
569	TyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyr	588
1657	CACCTCAGTGGAGTCCGAGTGTCTCAAAATCAGCTCTCATGACTCAGGTTTACGTTCT	1716
589	AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla	606
1717	GACAAACCAACCGCTCTCTCCATGATGCTGCAGACACAGCCCTATCTAACCCATATGCT	1776
607	IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIle	626
1777	CACGGCTCGGGCCATGTAGATCCCAAAAGGCTCTCTCACAGGTCTTGTCTACGACATC	1836
627	GlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeu	646
1837	TCAACCGAGGAATACATCAGGTTTGTGCTCTAGACTACACAGTCGATCATCATGTT	1896
647	AlaIleThrHisLys---AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPhe	665
1897	CGGATTGTGAAGCACTAGCGTTACTGCTGCTGAGGAAGTTC---TCAGATCTCGGT---	1950
666	SerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIle	685

```

Db 1951 CAGCTCAACTACCAAGTTTCTCGTTTGTGTTT---GGGGTAAAGAGTTGTGCGGTAC 2007
Qy 686 ThrArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAla 705
Db 2008 ACTCGGAAGTAACAAATGTTGTGTCAGCAAGCTCGGTTTACAAGTACGCGTTAATGGA 2067
Qy 706 ProGluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGln 725
Db 2068 GCTCTAGTGTGCGGAATCTCTGTTAAACCATCGAAACTTTCGTTTAAAGGCGTGGAGAG 2127
Qy 726 ThrLeuSerTyrArgValTyrPheValLeuLysLysLysAsnArgGlyGlyLysValAla 745
Db 2128 AAGAGAGGTACACAGTCACGTTGTTGTAGCAAGAAAGAGTGCAGTATCAACGAACAGGCT 2187
Qy 746 SerPheAlaGlnGlyGlnLeuThrTyrValAsnSerHisLeuMetGlnArgValArg 765
Db 2188 GAGTTT-----GGTTCATCATTGGAGCAATCCGAGCAC-----GAAGTGAGA 2232
Qy 766 SerProIleSer 769
Db 2233 AGTCCCGTTGCA 2244

RESULT 2
US-09-938-842A-1065
; Sequence 1065, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1065
; LENGTH: 2265
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1065

Alignment Scores:
Pred. No.: 3,13e-198 Length: 2265
Score: 1742.00 Matches: 370
Percent Similarity: 63.48% Conservative: 115
Best Local Similarity: 48.43% Mismatches: 241
Query Match: 43.35% Indels: 38
DB: 12 Gaps: 15

US-09-806-767-2 (1-775) x US-09-938-842A-1065 (1-2265)

Qy 11 IlePheLeuLeuPheCysSerSerSerSerGluIleLeuGlnLysGlnThrTyrIleVal 30
Db 52 CTCCTCTCTCTCTCCACACACACGCGC-----AAAAAACCTCATCATC 96
Qy 31 GlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTyrPheHisLeu 50
Db 97 CGGTCATAT---CACTCCGATTAACCGGAATCATCTCTACTCCACGATTTGGTACACA 153
Qy 51 SerPheLeuGlnGlnAlaValLeuGlyValGluGluGluGluGluProSerSerArg 70
Db 154 TCTCAACTCAATTCA-----GAATCATCT 177
Qy 71 LeuLeuTyrSerTyrGlySerAlaIleLeuGluGlyPheAlaGlnLeuThrGluSerGlu 90

```

```

Db 178 CTTCTCTACACTTACACACCTCTCTCCATGGCTTCTCCGCTTACCTCGACTCCACCGAA 237
Qy 91 AlaGlu---IleLeuArgTyrSerProGluValValAlaValArgProAspHisValLeu 109
Db 238 GCGGATCTCTCTCTCTCTCTCAATCAATCTCGATATCTCTCGAAGATCTCTCTCTAC 297
Qy 110 GlnValGlnThrTyrSerTyrLysPheLeuGlyLeuAspGly---PheGlyAsnSer 128
Db 298 ACACCTTACACATACGCTACTCTCGATTTCTCGGCTCTCAATCCGAATTCGGTGTTCAC 357
Qy 129 GlyValTyrSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGly 148
Db 358 GATCTCGGTTCTCTCTCT-----AACGCGGTTATCATCGGAGTTTATGATACTGCG 408
Qy 149 ValTyrProGluSerProSerPheAspAspThrGlyMetProSerIleProArgLysTyr 168
Db 409 GTTTGGCCGTAATCTAGAGCTTCGATGATCTGATGATGATGATGATGATGATGATGATG 468
Qy 169 LysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLysLeuIle 188
Db 469 AAAGGAGAAATGTGAATCTGGTTCCGATTCGATTCGAAGTTGTGTAAACAAGAGCTTATC 528
Qy 189 GlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsn 208
Db 529 GGAGCTAGAGAGCTTCTCCAAAGGATTCAAATGGCTTCTGTGTGTGTGTGTGTGTGTGT 588
Qy 209 MetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSer 228
Db 589 -----CGTGAATCTGTTCTCTCTCGTATGTTGACGGACATGGAACACATCTTCAACT 642
Qy 229 ThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyValAla 248
Db 643 ACCGCGGGGATCCCGGTTAGAAACCGTAGCTCTCTCGTTACGCGCGCGGTACGGCC 702
Qy 249 ArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTyrPheAsnGlyCys 268
Db 703 AGAGGTATGCCACCTCGTCTCGTCTCTACTTATAAGTTTGTGTGTGTGTGTGTGTGT 762
Qy 269 TyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspValLeu 288
Db 763 TTTGGATCTGATATACTAGCTCTATGATGATGATGATGATGATGATGATGATGATGAT 822
Qy 289 SerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThr 308
Db 823 TCGTTATCTCTGTGGTGGTCTCTCGTATTCGCGATACGATTGCGATTGAGCG 882
Qy 309 PheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIle 328
Db 883 TTTTCGGCTATGGAGAGAGGTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 942
Qy 329 GluSerSerValAlaAsnThrAlaProTyrValSerThrIleGlyAlaGlyThrLeuAsp 348
Db 943 AGAGCTTCTGTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1002
Qy 349 ArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeu 368
Db 1003 ACAGATTTTTCGGCTTTTTCGGAATCTCGGTAAACGGGAAACGACTTACCGGTGTTTCG 1062
Qy 369 TyrProGlyLysGlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGly 388
Db 1063 TATACGCTGTAGGAATG-----GGGACGAGCGCGCTTGAATTTGGTTTATAATAAGGG 1116
Qy 389 GlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGly 408
Db 1117 AATAGTAGTTCGAGTAATCTTTGTTTACTCTGGTTCGTTGATTCGAGTATTGTTCTGGG 1176
Qy 409 LysMetValIleCysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLys 428
Db 1177 AAGATTGTTGTTGTAGAGGTGTTAATGCTAGAGTTGAGAAAGAGGCTGTGGTTAGA 1236
Qy 429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluLysSer 448

```


Db 1237 GATGCTGGTGGTTAGGAGTATAATGGCGAATACCTGCTCGAGTGGAGGAGCTTG 1296
Qy 449 IleasValHisLeuLeuProAlaThrLeuLeuGlyThrGluSerValLeuLeuLys 468
Db 1297 GCGATAGTCAATGCTCCCGGATCGCTGTAGGGAAGAAGACTGGTATTACTTAG 1356
Qy 469 AlaTyrValAsnAlaThrValLysProLysAlaArgIlePheGlyThrValIle 488
Db 1357 GAGTATGTTAAGTCAGATTCTAAACACCGCTCTCTGTTTAAAGAACGGTCTT 1416
Qy 489 GlyArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAla 508
Db 1417 GACGTTAAGCGCTCTCTGCTGCTGCTTTAGCTCGAGAGTCTTAATCTGTTACT 1476
Qy 509 ProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleAlaLysPro 528
Db 1477 CCTGAATCTGAGCGCTGATGTTATGGTCTCGAGTAAATATTTGGCTGGTCT 1536
Qy 529 GlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsnPheThrValMet 548
Db 1537 GACGCTATTGGTCTACTGCTCTTGACAGGACTCTAGGAGGACTCAGTTCACATC 1596
Qy 549 SerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuLeuArgSerAla 568
Db 1597 TCAGTACGTCATGTCATGCCACACATCAGTGGTTAGCGGCTCTTTGAAAGCAGCT 1656
Qy 569 TyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyr 588
Db 1657 CACCTCAGTGGAGTCCGAGTGTATCAATCAGCTCTCATGACTACAGCTTACGTTCT 1716
Qy 589 AspArgGlnGlyLysAlaIleLysAspGly-----AsnLysProAlaGlyValPheAla 606
Db 1717 GACAACACCAACGCTCTCCATGATGCTGCGACACAGCCTATCAACCATATGCT 1776
Qy 607 IleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIle 626
Db 1777 CAGGCTCGGGCCATGATAGTATCCCAAGGCTCTCTCACCAGTCTGTTGTCACGACATC 1836
Qy 627 GlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeu 646
Db 1837 TCACCCGAGGAATACATCAGGTTTGTGCTCTAGACTACACAGCTCATCATGTT 1896
Qy 647 AlaIleThrHisLys-----AsnValSerCysAsnGlyIleLeuArgLysAsnProGlyPhe 665
Db 1897 GCGATTGTGAAGCAGCTAGCTTAACTGCTCGAAGAGTTC---TCAGATCTCTGTT--- 1950
Qy 666 SerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIle 685
Db 1951 CAGCTCAACTACCAAGTTCTCGGTTTGTGTT---GGGGTAAAGAGTGTGCGGTAC 2007
Qy 686 ThrArgArgValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAla 705
Db 2008 ACTCGGAAGTAAACAATGTTGTGCGCAAGCTCGGTTTACAAAGTGAAGGTTAATGGA 2067
Qy 706 ProGluGlyIleLysValIleValAsnProLysArgLeuValPheLysHisValAspGln 725
Db 2068 GCTCCTAGTGTGGAATCTCTGTTAAACCATCGAAACTTTCGTTTAAAGCGTGGAGAG 2127
Qy 726 ThrLeuSerTyrArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAla 745
Db 2128 AAGAAGAGGTACACAGTCACTGTTGTTAGCAAGAAAGAGTGTAGTATGACGAACAGGCT 2187
Qy 746 SerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArg 765
Db 2188 GAGTTT-----GGTTCATCCTTGGAGCAATCCGAGCAC-----GAAGTGAGA 2232
Qy 766 SerProIleSer 769
Db 2233 AGTCCCGTTGCA 2244

RESULT 3

US-09-938-842A-1186

; Sequence 1186, Application US/09938842A

; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kresps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1186
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-1186

Alignment Scores:

Pred. No.:	1,02e-170	Length:	2295
Score:	1589.00	Matches:	342
Percent Similarity:	60.36%	Conservative:	127
Best Local Similarity:	44.02%	Mismatches:	268
Query Match:	39.55%	Indels:	40
DB:	10	Gaps:	16

US-09-806-767-2 (1-775) x US-09-938-842A-1186 (1-2295)

Qy 7 PheLeuCysIleIlePheLeuLeuPheCysSerSerSerGluIleLeuGlnLysGln 26
Db 31 TTCCTCTCTTTCCGCTTATCTCTTCGACGCTCTCAGGCC-----GCGAAG 78
Qy 27 ThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysThrPheAlaSerLysPhe 46
Db 79 ACTTTCATTTTCCGTATCGATGATGCT---ATGCTTCTATTTCCGACGACACTAC 135
Qy 47 AspTrpHisLeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGlu 66
Db 136 CATGCTAT-----AGCACCGAGTTCGCGAAGAA--- 165
Qy 67 ProSerSerArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaAlaGlnLeu 86
Db 166 -----TCTCGAATCGTCCATGTTTACACACAGCTTCCATGCTTCTCCGCGCTTGT 219
Qy 87 ThrGluSerGluAlaGluIleLeuArgTyrSerProGluValValAlaValArgProAsp 106
Db 220 ACTCCAGATGAAGCAGATAATCTCCGTAAACCCACGAGTCTTGTCTGTTTCGAAGAC 279
Qy 107 HisValLeuGlnValGlnThrThrTyrSerTyrLysPheLeuGlyLeuAspGlyPheGly 126
Db 280 CGAGCTCGAGAGCTTCACACACAGCTTCTCTCAATTTCTTGTGTACAA-----AAC 333
Qy 127 AsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAsp 146
Db 334 CAAAGAAGACTATGTCAGATCTGATTCGATCAGACGATCAGATCAATCATTTGGCTTTT 393
Qy 147 ThrGlyValTrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArg 166
Db 394 ACCGGAATTTGGCGGAGCGGAGGAGTTCTCAGATCTTAACCTCGGTCCCAATCCAAA 453
Qy 167 LysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLys 186
Db 454 AGGTGGAGAGCGGCTTTCGCAATCCGAGCCAGATTCAGTCTCGGAATGTAACCGCTAAA 513
Qy 187 LeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAlaAsnSerProGluGluSer 206
Db 514 ATTATCGGAGCAAGATTCTTCGTAAAGGACAAACAGCGCT-----GTAATCGGAGGA 567

QY 207 ProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThr 226
Db ATCAACAAAACCGGTGAGTTCTATCTCCGTGACGCGGATGGACACGGTACTCACT 627
QY 227 AlaserThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGly 246
Db TCCTCAACCCGCGCGTCCGCTACGCTTTAAAGCGAGTATGTCGCGTTACGCGCTCCGCT 687
QY 247 ValAlaArgGlyMetAlaProGlyAlaHisIleAlaValIlyrIlyValCysTrp---Phe 265
Db GTAGCCAAAGGTGTGCTCCAAAAGCTCGTATCGCGCGCTACAAAGTCTGTTGGAAGAT 747
QY 266 AsnGlyCysTyrSerSerAspIleLeuAlaAlaIleAspValAlaIleGlnAspIlyVal 285
Db TCCGGTGTCTCGATTCCGATTCTCCGCGCTTCGATGCCGCTGTAGACGCGGTGC 807
QY 286 AspValLeuSerLeuSerLeuGly-----GlyPheProIleProLeuTyrAspAsp 302
Db GAGCTTATATCGATCTCAATCGGTGGTGGAGCGGATTAATTCGCGGTATTAACCTCGAT 867
QY 303 ThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAla 322
Db CCAATCGCTATAGCTGTACGCGCGCGCGCTGAAAGGAATCTTCGCTCTCTCTCTGCC 927
QY 323 GlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProIleValSerThrIle 342
Db GGAACCAAGGACCTACCGGTATGTCTAGTTACGAACCTCGCGCGGTGGTAAACCGGT 987
QY 343 GlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAlaAsnGlyIlyLeu 362
Db GGTCTAGTACATCGATCGGAATTCGCCAGCGATGCTATTCGCGCGACGACATCGT 1047
QY 363 LeuTyrGlyGluSerLeuTyrProGlyIlyLeuIlyAsnAlaGlyValGluValGlu 382
Db CTCAGAGGAGTGTCTTACGTGGAGTACCTTTAAAC-----GTCGTATGTTTCCG 1101
QY 383 ValIleTyrVal---ThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeu 401
Db GTGTTTATCCCGTAAATCCGGAATGTCTATCATCAGCGTCTATGATGAGACACGCTT 1161
QY 402 ProArgGluGluLeuArgGlyLysMetValIleCysAspArgGlyValAsnGlyArgSer 421
Db GATCCGAAGCAAGTGAGGGGTAAATAGTAAATCTGCGATAGAGAAAGCAGTCCACCGTA 1221
QY 422 GluLysGlyGluAlaValLysGluAlaGlyValAlaMetIleLeuAlaAsnThrGlu 441
Db GCCAAAGGATTGTTGTGAAGAAAGCAGGTGGTGTGCGAATGATCTCGTAAATGAGCA 1281
QY 442 IleAsnGlnGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyr 461
Db TCTAACCGTGAAGGATTAGTCGAGATGCTCATCTTATCCAGCCTGTCGCGTGGATCA 1341
QY 462 ThrGluSerValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIle 481
Db AACGAAGGAGATAGAAATCAAGCATATGCTCTTCATCCGATCCGATTCGTTCAATT 1401
QY 482 IlePheGlyThrValIleGlyArgSerArgAlaProGluValAlaGlnPheSerAla 501
Db GATTTCAGAGCACTATAGTTGGATTAAACCGGCTCCGCTATTGCTCTCTCTCCGCT 1461
QY 502 ArgGlyProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyVal 521
Db AGAGGACCAACCGGTTTAAAGCCCGGAGATCTTTAAACCGGATTGATTGCTCCCGGATT 1521
QY 522 AsnIleIleAlaAlaTyrProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArg 541
Db AACATCTCCCGCATGACAGACGCTGTTGGACCTACAGGTTTGGCGTCAGATCCAAG 1581
QY 542 ArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIle 561
Db AAAACGGGATTCACATCTCTCCGTTACTTCAATGGCATGTCTCTCAGCTTACTGGTGG 1641

QY 562 ThrAlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeu 581
Db CGCGCGCTTCTCAATCCGCTATCCGATTCGAGCCCTGCCGTGATACGATCCGCAATG 1701
QY 582 MetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAspGlyAsn----- 599
Db ATGACACGACTACCTACCTCGTGAATACTTAACCGCTCGTTGATCGATGAATCCACCGG 1761
QY 600 LysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn 619
Db AAATCGGCTACGCTTATGATTACGGTTCGGTCAATTAATTTGGCCCGCGCTATGAAT 1821
QY 620 ProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGly 639
Db CCGGCTCTGCTACGATATAACTTAATGATTAATACGTTCTTCTTCTTCTTCTTCTTCT 1881
QY 640 PheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeu 659
Db TAGGACCAAGACGATCCAGTGATACAAAGAACACCGGTGAGATGTCCGACGACGAGG 1941
QY 660 ArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePhe----- 676
Db AAACCGTCTCCGCGG---AATTTGAATTAATCTTCGATCAGCGCGGTGTTCTCTACTAAT 1998
QY 677 LysArgGlyLysThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsn 696
Db AGAAGAGATTGTCAGTAAACCTGTTATAAGGACGCGCGAGATGTGCGGACGCTGAG 2058
QY 697 SerIleTyrSerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLys 716
Db GCGGTTATCCGCGAGGATAGTCCGAGAGAGAGTACGCGTGACAGTGAACACCACT 2118
QY 717 ArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTyrPheValLeuLys 736
Db AGCGTTGTTTACTTCGCGCTTAAGACGCGGAGCTATGCGGTACAGTACGCTTAAT 2178
QY 737 LysLysAsn-----ArgGlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrp 754
Db ACAAGGATGTTGTTGGAGAAACAGGTCTCTGTTT-----GGTCACTCAGTCAGTGG 2232
QY 755 ValAsnSerHisAsnLeuMetGlnArgValArgSerProLysSerValThr 771
Db TTTGATGTGGAGAAACAGCTG-----GTTCCGAGCCCATCGTGGTGACC 2277

RESULT 4
US-09-938-842A-1186
; Sequence 1186, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1186
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1186

Alignment Scores: 1.02e-170 Length: 2295
Pred. No.:

Score:	1589.00	Matches:	342
Percent Similarity:	60.36%	Conservative:	127
Best Local Similarity:	44.02%	Mismatches:	268
Query Match:	39.55%	Indels:	40
DB:	12	Gaps:	16

US-09-806-767-2 (1-775) x US-09-938-842A-1186 (1-2295)

Qy	7	PheLeuCysAlrIlePheLeuLeuPheCysSerSerSerGluIleLeuGlnLysGln	26
Db	31	TTCCTCTCTTTCCGGTTTATCTTTCCAGCTCTTCAGGCC-----GCGAAG	78
Qy	27	ThrTyrlleValGlnLeuHisProAsnSerGluThrAlalysThrPheAlaSerLysPhe	46
Db	79	ACTTTCATTTCOGTATCGATGGTGATCT---ATGCCTTCTATTTTTCCGACGCCATAC	135
Qy	47	AsnTrpHisIleuSerPheLeuGlnGluLaValLeuGlyValGluGluGluGluGlu	66
Db	136	CATTGTAT-----AGCACCGATTCGCCGAGAA---165	
Qy	67	ProSerSerArgLeuLeuTyrrSerTyrglySerAlalleGluGlyPheAlaLaGlnLeu	96
Db	166	-----TCTCGAATCGTCATGTTTACCACACAGCTTCCATGCTTCTCCGCGTGT	219
Qy	87	ThrGlusSerGluAlaGluIleLeuArgTyrrSerProGluValValAlaValArgProAsp	106
Db	220	ACTCCAGATGAACAGCATTAATCTCGTAACCAACCACGAGTCTTGTGTTTTCSAAGAC	279
Qy	107	HisValLeuGlnValGlnThrThrrTyrrSerTyrllysPheLeuGlyLeuAspGlyPheGly	126
Db	280	CGAGTCGAGAGCTTCACACCACAGTCTCTCTCATTTCTTGTTACAA-----AAC	333
Qy	127	AsnSerGlyValTrpSerLysSerArgPheGlyGlnGlyThrIlelleGlyValLeuAsp	146
Db	334	CAAAAAGGACTATGTCAGATACTGATACGGATCAGACGTAATCATTTGCGTTTTGAC	393
Qy	147	ThrGlyValTrpProGluSerProSerPheAspAspThrGlyMetProSerIleProArg	166
Db	394	ACCGGAATTTGGCCGAGCGGAGGAGTTCTCAGATCTTAACCTCGGTCCAATTCAAAA	453
Qy	167	LysTrpLysgLyileCysGlnGluGlyGlusterPheSerSerSerCysAsnArgLys	186
Db	454	AGGTGGAGAGCGCTTTGCGAATCCGAGCCAGATTTCAGTCTCGGAATGTAACCGTAA	513
Qy	187	LeulleGlyAlaArgPhePheilleArgGlyHisArgValAlaAsnSerProGluGluser	206
Db	514	ATTATCGGACAGATTCCTCGTAAGGACACAACAGCCGT-----GTAATCGGAGA	567
Qy	207	ProAsnMetProArgGluTyrlleSerAlaArgAspSerThrGlyHisGlyThrHisThr	226
Db	568	ATCAACAAAACCGTTGAGTTTCTATCTCTCGTAGCCGCGATGACACGCTACTCACAC	627
Qy	227	AlaserThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGly	246
Db	628	TCCTTCACCCCGCTGGCCGTACCGCTTTTAAAGCGAGTATGTCGGGTACGCCCTCGGT	687
Qy	247	ValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrlLysValCysTrp---	265
Db	688	GTACCAAAGGTGTGCTCCAAAAGCTCGTATCGCGCGCTACAAGTCTGTGGAAAGAT	747
Qy	266	AsnGlyCysTyrrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysVal	285
Db	748	TCCGGTTGTCTCGATTCCGATATTCTCGCGCGCTTTGATGCGGTGTTTAGACACGGTGC	807
Qy	286	AspValLeuSerLeuSerLeuGly-----GlyPheProIleProLeuTyrrAspAsp	302
Db	808	GACGTTATATCGATCTCAATCGTGTGGAGACGGGATTTACTTCGCGGTATTACTCTCAT	867
Qy	303	ThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyVileSerValIleCysAlaLa	322
Db	868	CCAATCGCTATAGCTGTACGCGCGCGGTGMAAAGGANAICTTCGTCTCTCTCTCTGCC	927
Qy	323	GlyAsnAsnGlyProilledUserSerSerValAlaAsnThrAlaProTrpValSerIle	342

Db 1999 AGAAGAGGATTGGTGAATAAACTGTTATTAAGACCGCGCAAGAAATGTCGGCAGGCTGAG 2058
Qy 697 SerileTyrSerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLys 716
Db 2059 GCGGTTATCGGCGCAGATAGAGTCGCCGAGAGGAGTGACGGTGACAGTGAACCACT 2118
Qy 717 ArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTrpPheValLeuLys 736
Db 2119 AGGCTTGTTACTTCGGCGGTTAAGACGAGCGAGCTATCGGTTACAGTGACGGTTAAT 2178
Qy 737 LysLysAsn-----ArgGlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrp 754
Db 2179 ACAAGGAATGTGTTGGGAGAAACAGGTGCTGTGTT-----GGTCAGTCACGTGG 2232
Qy 755 ValAsnSerHisAsnLeuMetGlnArgValArgValArgSerProLysSerValThr 771
Db 2233 TTGATGCTGGGAAACAGTG-----GTTCCGAGGCCCAATCGTGTGACC 2277

RESULT 5
US-10-260-238-1206
; Sequence 1206, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 1206
; LENGTH: 2043
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (389)..(389)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-1206

Alignment Scores:
Pred. No.: 1.18e-136 Length: 2043
Score: 1290.50 Matches: 297
Percent Similarity: 55.20% Conservative: 112
Best Local Similarity: 40.08% Mismatches: 245
Query Match: 32.12% Indels: 87
DB: 12 Gaps: 17

US-09-806-767-2 (1-775) x US-10-260-238-1206 (1-2043)

Qy 50 LeuSerPheLeuGlnGluAlaValLeuGlyValGluGluGluGluGluProSerSer 69
Db 16 TTGCTCTTCTCTTGTGTTTCATCTCCTTCATATTACACA----- 54
Qy 70 ArgLeuLeuTyrSerTyrGlySerAlaIleGluGlyPheAlaGlnLeuThrGluSer 89
Db 55 -----ACTAGTTCACCGGTACTGAAATTTTCATGCACTCT----- 90

Qy 90 GluAlaGluIleLeuArgTyrSerProGluValValAlaValArgProAsp----- 106
Db 91 -----AGGCTAGACACATATATAGTGGCTGTGCTGCCACCAAACTTC 135
Qy 107 -----HisValLeuGlnValGln 112
Db 136 TCAATTGATGATGAGCAACATCAAGCTGAGAGTGGTACAGATCATTTCTCCCACTCTTG 195
Qy 113 ThrThrTyrSerTyrLysPheLeuLeuAspGlyPheGlyAsnSerGlyValTrpSer 132
Db 196 ACNAGCCACACACAGATTTCCTAGGCGTGGT-----CTCAGGGAAGGATCTCGAAA 249
Qy 133 LysSerArgPheGlyGlnGlyThrIleGlyValLeuAspThrGlyValTrpGlu 152
Db 250 AAAACTAGCATGGTGGGTTGTCATCATAGTGTCTCTTGACACAGGATGATTCACT 309
Qy 153 SerProSerPheAspAspThrGlyMetProSerIleProArgLysTrpLysGlyLeu 172
Db 310 CATACCTCATTTGATGATGATGGCATCGAGGACCCCACTAAATGCGCGGTCTCTGC 369
Qy 173 GlnGluGlyGluSerPheSerSerCysAsnArgLysLeuIleGlyAlaArgPhe 192
Db 370 AAGTCA-----TCTCTGATGAATNCAACAAGAACTCATTTGGGGTAGTCA 417
Qy 193 PheIleArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArgGlu 212
Db 418 TTCATTCGGGTCAGAAATCAGCA----- 441
Qy 213 TyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGly 232
Db 442 -----CCTCCACCGGATGATGATGGCATCGGACACATACCGCAAGCAGCGAGTGGC 495
Qy 233 SerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAla 252
Db 496 GGTGTTGTAGATGGTCAAGCGTGTGTCGCAATGCAATGCGACAGCAGCGTGGCATGGCT 555
Qy 253 ProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAsp 272
Db 556 CCTGTGTGCACACCTTGCTATCTACAGTGTGACGATAAAGTTGCGGTGATCTGAC 615
Qy 273 IleLeuAlaIleAlaAspValAlaIleGlnAspLysValAspValLeuSerLeuSer 292
Db 616 ATACTTGTGCGCATGGAGCGGCTATTGCTGATGGTGTGACATCATGTCTATCTCTT 675
Qy 293 GlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMet 312
Db 676 GGTGTCCACGCGAAACCATTTTACATGATATAATAGCCACTGTCATCTTTTCTGCCATG 735
Qy 313 GluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerVal 332
Db 736 AGGAAAGGGATCTTTGTAAGTCTTGTGCGAGAAATTTCTGTGTCATCTTCCAGCACTCTA 795
Qy 333 AlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPhePro 352
Db 796 AGCAACAGGCGCCATGGTTCTGCTGCGAGCAATTTCTGTGTCATCTTCCAGCAATGAA 855
Qy 353 AlaValValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLys 372
Db 856 GCCCTTGTCAAGCTAGGTGATGGAGACTTATTTCTGGCGAATCTGCTTATCAACCAT 915
Qy 373 GlyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGly 392
Db 916 AATCTTGTAT-----CCCTTAGAGTTAGTGTACCCACAG-----ACTTCCGGT 957
Qy 393 SerGluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIle 412
Db 958 CAATATTTCTCTCTT-----CTGAAGATGTTGCGAGAAAGATTGTTGCC 1005
Qy 413 CysAspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGly 432
Db 1006 TGTGAGCAC---ACAACCTCATCAGACATTATTGACGCTTCGTCAGGATGCTGGTGA 1062
Qy 433 ValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHis 452

```

; Sequence 485, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; PRIOR FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 485
; LENGTH: 2073
; TYPE: DNA
; ORGANISM: Oryza sativa
; US-10-260-238-485

Alignment Scores:
Pred. No.: 1,07e-132 Length: 2073
Score: 1256.00 Matches: 300
Percent Similarity: 58.16% Conservative: 117
Best Local Similarity: 41.84% Mismatches: 240
Query Match: 31.26% Indels: 60
DB: 12 Gaps: 23

US-09-806-767-2 (1-775) x US-10-260-238-485 (1-2073)

QY 86 LeuThrGluSerGluAlaGluLeuArgTyrSerProGluValValAlaValArgPro 105
DB 1 ATGACGAGCAGCAGGCGCGGCACATCCGACCATCCCTGCTGCTCCGCTATACCCC 60
QY 106 AspHisValLeuGlnValGlnThrThrTyrSerTyrIysPheLeu----- 120
DB 61 GACGAGCATCTCCAACTGCACACACTCAGTCTCCCTCCTCTCCTCCTCCTCCGCT 120
QY 121 ---GlyLeuAspGlyPheGlyAsnSerGlyValTrpSerIysSerArgPheGlyGlnGly 139
DB 121 GTTGGCTCGTACAGGCATCAACGCGGC-----GGCAGCGGC 159
QY 140 ThrIleGlyValLeuAspThrGlyValTrpProGluSer---ProSerPheAsp--- 157
DB 160 GCGGTGATCGCATCTTGACACTTGCACATTCATCCAGAGCGCGCAATCTCTTCC 219
QY 158 AspThrGlyMetProSerIleProArgIysTrpIysGlyIleCysGlnGluGlySer 177
DB 220 GACTCGTGTTCCTCCACCATCTCCGACCTTCGCTGCTGCTGCTGCTGCTGCTG 279
QY 178 PheSerSerSerSer---CysAsnArgIysLeuIleGlyAlaArgPhePheIleArgGly 196
DB 280 TTTATGCGCCGCTACTGCAACAACAGCTCGTCGCGGCAAAAGTTTCTATAAGGG 339
QY 197 His-----ArgValAlaAsnSerProGluGluSerProAsnMetProArg 211
DB 340 CACGAGGCCAGATGGTCACTGATCAATGAGACACAAGATCCCAAG----- 387
QY 212 GluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGly 231
DB 388 -----TCACCTCTCGACAGGCGCCAGCGCACGACACTGCTATCTACCCGCT 438

RESULT 6
US-10-260-238-485
```

```

1063 TCTGGTGTGATCTCTAGCTAGGAGGATAGTGGCCACATAACATTTGCTGACCCAAAT 1122
QY 453 LeuLeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLeuIysAlaTyrValAsn 472
DB 1123 GTCTACCGGTGCTCTATGTGGATTTCCAGATGCAACAGTATCAGACAGTACATTAAT 1182
QY 473 AlaThrValIysProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSerArg 492
DB 1183 TCAAGTAACAGTCCAACTGCTCAATCATCTTCAATGGTACCTCATTTGGGGAACCTCA 1242
QY 493 AlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeu 512
DB 1243 GCTCTGTAGTGTATTTCTCATCTCGTGGCCCTAGTACTCGGAGCCCGAGTATCTT 1302
QY 513 LysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGly 532
DB 1303 AAACCTGACATCATTTGCTCTGGAGTAAAGTCACTTCTGCTGCGGCTTCATGGAGGG 1362
QY 533 ProThrGlyLeuProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSer 552
DB 1363 CAAGACGCA-----AACAATGATAAACAACAGGACATTCAACTGCTTATCTGGAACATCA 1416
QY 553 MetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyrProAsnTrp 572
DB 1417 ATGTCAACTCTCACTTGTGGGCTGCTGCTCATCAAGGGAACACACCCAGATGG 1476
QY 573 SerProAlaIleIysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGly 592
DB 1477 TCATCCGAGCATCAAGTCCGCAATCATGACACAGCTTATGTTGTTGATAACGAGAAG 1536
QY 593 LysAlaIleIysAspGlyAsn---LysProAlaGlyValPheAlaIleGlyAlaGlyHis 611
DB 1537 AAGCAATCTGGATGAAGATACAATATTGCGAGGCAATTTCGAGTTGTTGTTGACAT 1596
QY 612 ValAsnProGlnLysAlaIleAsnProGlyLeuValIysAsnIleGlnProValAspTyr 631
DB 1597 GTTAGTCTCTGAAGCCATTTGACCCCGGTCTGATCTATGACATGATGATGACAAATAC 1656
QY 632 IleThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLys 651
DB 1657 ATTTCATATCTCTGTGGCTGGTTACACAGATGTTCAAGTTGAGTAATGCAATCAG 1716
QY 652 AsnValSerCysAsnGlyIleLeuArgIysAsnProGlyPheSerLeuAsnTyrProSer 671
DB 1717 AAGATGCTCAAGGA-----AGCAAAATACTGAAGTGAACCTTAATCTCATCA 1770
QY 672 IleAlaValIlePheLysArgGlyIysThrThrGluMetIleThrArgArgValThrAsn 691
DB 1771 GTAGCAGTAAGGCAAGTGTGCAAGCTT-----GTGGTGAACCGGACTGTTACTAAT 1824
QY 692 ValGlySerProAsnSerIleTyrSerValAsnValIysAlaProGluGlyIleLysVal 711
DB 1825 GTGGAGAGCAAAATTCAGCTACACCGTGAATATGATGCTGAGGAGGTGATGACA 1884
QY 712 IleValAsnProLysArgLeuValPheLysHisValAspGln-----ThrLeuSerTyr 729
DB 1885 AGTGTGTACCAACAAGCTAGATTTACCAAGATGAGGAGAGAGAAAATTTCTATTA 1944
QY 730 ArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAlaSerPheAlaGln 749
DB 1945 AGCTTAAGTTGGATATTAGCAAAACTAACCAT-----GCCGAG 1983
QY 750 GlyGlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSer 769
DB 1984 GGAAGCTTCAAGTGGGTATCTGAGAAGCATGTT-----GTGAGGAGCCCATTTGCA 2034
QY 770 Val 770
DB 2035 ATT 2037
```

Qy	232	GlySerValSerMetalAsnValLeuGlyAsnGlyAlaGlyValAlaAlaArgGlyMet	251
Db	439	GGTTCCGGGTCCCTGGCGCTAACTTTGGCTATGCCAATGCACCGCCCAAGGCATG	498
Qy	252	AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAsn-----Gly	267
Db	499	GCTATACGTGCACACATTCGATCTCAAAAGTATGCTGGAGAGATGATGGAAATGCATCG	558
Qy	268	CysTyrSerAspIleLeuAlaAlaIleAspValAlaIleGlnAspLysValAspVal	287
Db	559	TGCCTACCTTCACATCTTCGCGGAATGATGAGGCNAATTCAGATGGATAGATGTC	618
Qy	288	LeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGly	307
Db	619	ATATCCCTTTCTTGGGTGCTAAAGCCACAACTCTACAACGAAACCCACCTCCCTTGGC	678
Qy	308	ThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyPro	327
Db	679	GCATTCATAGCCATCCGCTAGAGGATAGTCGTCCACCTCCGCTGGAAATGATGGCCCT	738
Qy	328	IleGluSerSerValAlaAsnThrAlaProTrpValSerThrIleGlyAlaGlyThrLeu	347
Db	739	GGCAGCTACTGCTAATACTCGCACCATGGGTGATAACGGTTGGCGCATCAAGCAT	798
Qy	348	AspArgArgPheProAlaValValArgLeuAlaAsnGlyLysLeuLeuTyr---GlyGlu	366
Db	799	GACCGTCGGTTCCCGAGCTCACTGTTCTTGGACATAATCGCGAAACCTACATCGGCACC	858
Qy	367	SerLeuTyrProGlyLysGlyIleAsnAlaGlyArgGluValGluValIleTyrVal	386
Db	859	TCGCTGTATTTCGGCCAA-----AACACACCGGTTTCATTTCACCTCTAGTGTAC	909
Qy	387	ThrGlyGlyAspLysGlySerGluPheCysLeuArgGlySerLeuProArgGluGluIle	406
Db	910	---GGTGGTGAATGCTGTTCTGCTCTGTGAATACGGGATGCTTCCAGCAACATGGTC	966
Qy	407	ArgGlyLysMetValIleCys-----AspArgGlyValAsnGlyArgSerGlyLysGly	424
Db	967	ACCGAAAGATTGTCTCTGCTATGGCACAAAAAATACTACCAACCCCATAGTCAAGAA	1026
Qy	425	GluAlaValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGln	444
Db	1027	GCTGCTGTCCAAACAGCGGTGGTGGTGCCATCATTTCAATTGCTCCAGAAATATGGT	1086
Qy	445	GluGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyTyrThrGluSer	464
Db	1087	GATTTCTTACAGACCTCGCTGATATCTTGCACAGCTGCAGATTAACCTTCAAGGACACC	1146
Qy	465	ValLeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGly	484
Db	1147	GAACAGATCCATTGTCACACACACGTCGGTGGCCGACCGGTTGCAAGATCGACTTCTTA	1206
Qy	485	GlyThrValIleGlyArgSer---ArgAlaProGluValAlaGlnPheSerAlaArgGly	503
Db	1207	GGCACTGTGATCAACCAATCACCTGCTGCTCCAAAGGTTCGAGCGCTTCTCAAAGTCGTGA	1266
Qy	504	ProSerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsnIle	523
Db	1267	CCGAACCGTTTGTCTCAGAGATCCTCAAGCCGACATGATGTCGCCCGGTGTGCAATC	1326
Qy	524	IleAlaAlaTrpProGlnAsnLeuGlyProThr---GlyLeuProTyrAspSerArgArg	542
Db	1327	CTCGCTCATGGACCGGAGAAATCTCGCTACCATGGCCCATGTGCATGCACAATAGCGGC	1386
Qy	543	ValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThr	562
Db	1387	GTCAGTTCAACATCATCTCCGGCACGTCCATGGCGTGCCTGTCATCATGAGCGCATGCC	1446
Qy	563	AlaLeuIleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMet	582
Db	1447	GCTATGCTTAAAGTGGCACAAACCGAGTTGGAGCCCGGCAATCAAGTCGGCCATGATG	1506
Qy	583	ThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIleLysAsp-----GlyAsnLys	600

Db	1507	ACGACGCCCTACAAATGTGGACAATGATGGCAATGCGCATCAAGACACATGGCCACGGGCCAG	1566	...
Qy	601	ProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnPro	620	
Db	1567	GCAGCCAGGCGGTCCAGCTTGCGTCTGGCCACGTCGACCCCAACGCGCCCTCGACCCCT	1626	...
Qy	621	GlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeuGlyPhe	640	
Db	1627	GGATTGGTGAACAACACCACCGCGGAGGATTACATAACATCTCTGCAGCCTCGGCTAC	1686	...
Qy	641	ThrArgSerAspIleLeuAlaIleThrHis-----LysAsnValSerCysAsnGlyIle	658	...
Db	1687	AACAGTATCAGATCGCGCTCTCACAAATGACGGTTCGACGACAGACTGCTCGACCAGG	1746	
Qy	659	LeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArg	678	
Db	1747	CCACGGAGGTCGGTGGCC--GACCTCAACTACCCGGCGTTCTCGGTGGTGTTCGTGAGG	1803	
Qy	679	GlyLysThrThrGluMetIleThrArgArg-----ValThrAsnValGly---SerPro	695	...
Db	1804	-----TCCGGAGACAGGTCAGCGACGCGCGCGGTGACCAACGTCGGGGGGAACACC	1857	...
Qy	696	AsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLysValIleValAsnPro	715	
Db	1858	AATGTCATGACAACTTACATCACCGCACCCCGCCACCACTCTCACGGTGACGCCA	1917	...
Qy	716	LysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTyrPheValLeu	735	...
Db	1918	ACGAGGCTCCGTTCCAGCGCGGACGAGAACGCTGGATTACTCCATC-----	1965	...
Qy	736	LysLysLysAsnArgGlyGlyLysValAlaSerPheAlaGln---GlyGlnLeuThrTrp	754	...
Db	1966	---ACCGTGTCCGAGGGCGCACCGACTCGCTCCGACGACCAACGAGTGGGATCGATCGTCTGG	2022	
Qy	755	ValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerValThr	771	
Db	2023	AGGCATGGCCAAACAT-----ACGTTGACGAGCCCGCTCGTGGGCACC	2064	...

RESULTS

```

RES001 /
US-10-259-165-271
; Sequence 271, Application US/10259165
; Publication No. US2003013588A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: Patentlist.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 271
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-165-271

```

Alignment Scores:	2,19e-123	Length:	2205
Pred. No.:	1175.00	Matches:	293
Score:	52.87%	Conservative:	252
Percent Similarity:	37.32%	Mismatches:	122
Best Local Similarity:	29.24%	Indels:	118
Query Match:	13	Gaps:	27
DB:			
US-09-806-767-2 (1-775) x US-10-259-165-271 (1-2205)			
Qy	38	ThrAlaLysThrPheAlaSerLysPheAspTrpHisLeu---SerPheLeuGlnGluAla	156
Db	10	ACTCTATCTTCATTCCGCACACACGTTGACTGGGTCAATATCATCATTTCCCTCTCATC	69
Qy	57	ValLeuGlyValIuGluGluGluGluGluProSer---	68
Db	70	GTCTACATGGGGAGAGAACATGATCGTCTGTGTACGGCATCGCATCATGAC	129
Qy	69	-----SerArgLeuLeuTySerTyr	75
Db	130	ACGCTTACTTCTGTTCTGGAGTAAGATGGCGCCATGAAGTCTTAGTGTCAGATTAC	189
Qy	76	GlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluIleLeuArg	95
Db	190	AAACATGGGTTCTCTGGATTTCGCGGATGCTCACCGAATCCAGCTGAGGAACITGCA	249
Qy	96	TyrSerProGluValValAlaValArgProAspHisValLeuGlnValGlnThrThrTyr	115
Db	250	AGATTACCTGAAGTCAATCAGCGTGAAGCTAACACTTACCATCAAGCACAAAGCATCGA	309
Qy	116	SerTyrLysPheLeuGlyLeuAspGlyPheGlyAenSerGlyValTrpSerLysSerArg	135
Db	310	AGCTGGGATTTCTCGGCTTAACATACACGAACATCAGGTCTTCTTAAAGAAAGCAAG	369
Qy	136	PheGlyIleGlnGlyThrIleIleGlyValLeuAspThrGlyValTrpProGluSerProSer	155
Db	370	AACGGTCAAGATGTCATTCTCGGTGTCATCGATTCCAGGTATATGSCCGGAATCAGGAAC	429
Qy	156	PheAspAspThrGlyMetProSerIleProArgLysTrpLysGlyIleCysGlnGluGly	175
Db	430	TTTCGATGACAAATGGCTTACAGTCCCGTGCCGACGATGGAAGGGCAAGTGCACACCGGC	499
Qy	176	GluSerPheSerSerSerSer---CysAenArgLysLeuIleGlyAlaArgPhePheIle	194
Db	490	GGCGCGTTTCAACGCCACACCGGTTGCACAGGAAGATCATCGCGTACGATGATCTCG	549
Qy	195	ArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProA-GgiuTyrIle	214
Db	550	GGCGGC-----ATTCGACGACGAG-----AACCTGAAGGGCGAGTACATG	598
Qy	215	SerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSer	234
Db	589	TCGGCAGGAGACCTCGCGCGCCACGCGACGACGTCGCGCTCGACGATCTGTCGGCGGCGAG	648
Qy	235	ValSerMetAlaAsnValLeuGlyAenGly-----AlaGlyValAlaArgGlyMet	251
Db	649	GTACGGAACGTGAGCCACACAGAGGGCGGCGCCCTGGCTGCGGACAGGGCGCGGTGGG	708
Qy	252	AlaProGlyAlaHisIleAlaValTyrLysValCysTrpPheAenGly-----	267
Db	709	GCGCCGCGCGCGGCGTCTGCTTCTACAGGTGTCTGG-----GGCCTGCGCGCCGAC	762
Qy	268	CysTyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLysValAspVal	287
Db	763	TGTGGGGAGCGCGATCTCCGCCCATTCGACGACGCCCATGAACAGCGCGGTGGATGTG	822
Qy	288	LeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIleAlaIleGly	307
Db	823	CTGTCTCTGTCATAGAGCGCGGGC---GAGCATTTACGAG-----	861
Qy	308	ThrPheArgAlaMetGluArgGlyIleSerValIleCysAlaAlaGlyAenAsnGlyPro	327

```
Qy 669 TyrProSerIleAlaValIlePheLysArgGlyLysThrThrGluMetIleThrArgArg 688
Db 1882 CTCCTGGTCAATTCAGTGCCTAAAGACTCTGTCCAG-----GTTTGGCGGACT 1935
Qy 589 ValThrAsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGly 708
Db 1936 GTCACCAATGTTGAGAGCAGAGCAACTTACCATGCACCGTGTGAGGCACCGGTGGT 1995
Qy 709 IleLysValIleValAsnProLysArgLysValPhe---LysHisValAspGlnThrLeu 727
Db 1996 ATGACTATGTCAGTGGACCTTCGGTGCATCACATTCATAGGCGGTGTAGTAAAGTGA 2055
Qy 728 SerTyrArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAlaLysPhe 747
Db 2056 ACATTTAAGGTGACATTCACACACACAGAGAGTACAAAGTGGC-----TAC 2103
Qy 748 AlaGlnGlyGlnLeuThrTrpVal-----AsnSerHisAsnLeuMetGlnArgValArg 765
Db 2104 ACGTTTGGAGTTTGACATGCTAGATGTTAACAACACACTCT-----GTGAGA 2151
Qy 766 SerProIleSerVal 770
Db 2152 ATTCGATTGCAGTT 2166
```

RESULT 8

```
US-10-260-238-1133
; Sequence 1133, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 1133
; LENGTH: 2334
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (498)..(498)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N region
; LOCATION: (548)..(548)
; OTHER INFORMATION: n = any nucleotide
; FEATURE:
; NAME/KEY: N region
; LOCATION: (551)..(551)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-1133
Alignment Scores: 2.12e-121 Length: 2334
Pred. No.: 1158.00 Matches: 291
Score:
```

```
Percent Similarity: 50.73% Conservative: 127
Best Local Similarity: 35.32% Mismatches: 256
Query Match: 28.82% Indels: 150
DB: 12 Gaps: 26
US-09-806-767-2 (1-775) x US-10-260-238-1133 (1-2334)
Qy 16 CysSerSerSerSerGluIleLeuGlnLys----- 25
Db 61 TGTAGTAGAGCAGGAGCGTGTCTCGTACTCCGAATGGCAGCGCGCGCGCGCGCGCGCG 120
Qy 26 -----GlnThrTyrIleValGlnLeuHisProAsnSerGluThrAlaLysPheAla 43
Db 121 GCCTACCGGACTTACATCTGCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 174
Qy 44 SerLysPheAsp-----TrpHisLysSerPheLeuGlnGluAlaVal 57
Db 175 -----GACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 219
Qy 58 LeuGlyValGluGluGluGluGluProSerSerArgLeuLeuTyrSerTyrGlySer 77
Db 220 ---GGCGCGCATGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 276
Qy 78 AlaIleGluGlyPheAlaGlnLeuThrGluSerGlu---AlaGluIleLeuArgTyr 96
Db 277 GTGTTGAGAGGCTTCGCGCGAGGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336
Qy 97 SerProGluValValAlaValAlaValArgProAsp---HisValLeuGlnValGlnThrTyr 115
Db 337 AAGCCCGGGTTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 396
Qy 116 SerTyrLysPheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrp---SerLysSer 134
Db 397 AGCGCGGAGTTCCTCGGCGTACG-----AGAGCGCGCGGTTCTGGAGGAGCGTCGCC 450
Qy 135 ArgPheGlyGlnGlyThrIleLeuValLeuAspThrGlyValTrpProGluSerPro 154
Db 451 GGTACGAGGAGGAGGCGTGTGCGGCGTCTGCACACCGCGCGTCCANGCGCGCGCGCT 510
Qy 155 SerPheAspAspThrGlyMetProSerIleProArgLysTrpLysGlyLeuCysGlnGlu 174
Db 511 TCCTTCGACGACCGCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570
Qy 175 GlyGluSerPheSerSerSerSerCysAsnArgLysLeuIleGlyAlaArgPheIle 194
Db 571 GCC-----GCCACCGCGCGGTGCAACAAAGCTCGTCGGCGTCAAGTCTCTCGTT 621
Qy 195 ArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArgLysTyrIle 214
Db 622 GACGCGCGCGCG----- 633
Qy 215 SerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerSer 234
Db 634 GCGCGCGACGACGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
Qy 235 ValSer---MetAlaAsnValLeuGlyAsnGlyValAlaArgGlyMetAlaPro 253
Db 694 GTCGCGCGCGCGCGTCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
Qy 254 GlyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAspIle 273
Db 754 GCGCGCGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
Qy 274 LeuAlaAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGly 293
Db 814 CTAGCGCGGTTTCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
Qy 294 GlyPheProIleProLeuTyrAsp---AspThrIleAlaIleGlyThrPheArgAlaMet 312
Db 874 AGGTGTCGAGTCCCGCTTCGATGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
Qy 313 GluArgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerVal 332
```


QY 221 GlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnVal 240
 Db 665 GGCCATGGCTCCACACTCTCTCCACCGCGCGGAGATTTCCTCTGGAGTCAGCATC 724
 QY 241 LeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyr 260
 Db 725 TTCGGCCAAAGCAACGCGACCGGCAAAAGGTGGTCCCTCCGCCACGAGTGGCGCGGTAC 784
 QY 261 LysValCysTrp-----PheAsnGlyCysTyrSerSerAspIleLeuAlaAla 276
 Db 785 AAAGTCCTGGCGCCAGTCAGTCAAGGCAACGAATGCTACGACGACGTCCTCGTGCT 844
 QY 277 IleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPhePro 296
 Db 845 TTCGACGCTGCCATACACGACGGAGCTGAGTCAATCCGCTCTCTAGCGCGGAGCCT 904
 QY 297 IleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIle 316
 Db 905 ACCTCGTGTTCCTAACGACAGTGTGCTGATTGGTTCCTTTCACGCGCTAAGAAGCGGATC 964
 QY 317 SerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAla 336
 Db 965 GTGGTGTTCCTCTCGCGGAACCTCTGGACCGCGGAGATAGTACGGTTCCTAATGTCGG 1024
 QY 337 ProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValAlaArg 356
 Db 1025 CCATGGCAGATTACCGTGTGTGTAGCACCATGGACCGGTGAGTTTGTAGCATCTTGT 1084
 QY 357 LeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsn 376
 Db 1085 CTCGGTAATGGAACA-TTACAAGGTTAGATATTAGTCCAAGATATGGATCATATCT 1143
 QY 377 AlaGlyArgGluValGluVal-----IleTyrValThrGlyGly 389
 Db 1144 CTATCTCAAACTCTGAAATGAATCTTGAAATGAGCTTGAATCTTTTCATTCGTAGG 1203
 QY 390 AspLysGly----- 392
 Db 1204 GACAAAGCTGTCTATCCACGGCATGGCCATGCTAAGTCTATCCAAATATGGCATCTG 1263
 QY 392 ----- 392
 Db 1264 TAAATGCAAAAGCTAAGAACGCTTCAGCTTTGGATGGGTGAGTTGGCTAATCTCCTTAC 1323
 QY 393 -----Ser 393
 Db 1324 TTAATTACGCCAAATGTTTATGTTAATGTTAATTTGAATCTCTCTCTTGTGTAGA 1383
 QY 394 GluPheCysLeuArgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCys 413
 Db 1384 CAATTATGCAAACTGGATCGCTTGACCTATATAAGACAAGGGAAGATATTAGTGTGT 1443
 QY 414 AspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyGlyVal 433
 Db 1444 CTCAGAGGCGAAGACGGAGGGTGGAAAAGGTTCGGCCGTGGCTTTAGGGGTGGCATA 1503
 QY 434 AlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeu 453
 Db 1504 GGGATGTTCTTGGAACACCTTATGTCACCGGGAATGATTACTCGCTGACCGCATGC 1563
 QY 454 LeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAla 473
 Db 1564 CTTCCAGCCACACAGCTCACTTCTAAGGATAGTTTGTCTGTGTCAAGATATATCAGCCAA 1623
 QY 474 ThrVal----- 475
 Db 1624 ACCAAGTAGGATTCCTGATTACACTTTACCTAAACTGTTTCAAGTTCGAAGAAAC 1683
 QY 476 -----LysProLysAlaArgIleIlePheGlyThrValIleGly 489
 Db 1684 TTTATTTTGTAGTAGAAGCGGATTGGACATTAATCTCTCTCAAGGACAGATTGGGA 1743

QY 490 ArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnPro 509
 Db 1744 CTAATAACAGCGCCTGTAATGGCTTCATTTTCTCAAAGGTCCAAGCATAGTGGCTCT 1803
 QY 510 SerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGln 529
 Db 1804 CAGATTCTGAAGCCCTGACATACTGCTCCTGGTGTGAGTGTGATCGCTCCCTACACTGA 1863
 QY 530 AsnLeuGlyProThrGlyLeuProTyrAspSerArgArgValAsnPheThrValMetSer 549
 Db 1864 CGAGTCTCTCCAAACAAACGACAGTTGATCTCTCGAGACTTCTGTTCAATGCTATTCA 1923
 QY 550 GlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyr 569
 Db 1924 GGAACCTCTATGCTCTGTCTCATATCTCTGSCATTCGGGTCTTCTCAAAACTCGTTAT 1983
 QY 570 ProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAsp----- 586
 Db 1984 CTTCTCTGGACCGCGCAGCTATCCGCTCTCCCATCATGACTACCGGTAAAGATGTATTT 2043
 QY 586 ----- 586
 Db 2044 CGATTCTAGAGCTTACACATTGTTGACAAAATTGAAGTTTAATTCTATTATTACCACTAA 2103
 QY 586 ----- 586
 Db 2104 AACACATTAAACATCTACTGGTCTGGACTTATAACAAATCTGTGTTCTTCCATGGCCAT 2163
 QY 586 ----- 586
 Db 2164 GTTCTGTTTTCTTAGTTTGGAAATATTGGAATAGAGTTTCTTGAATCTTACAGT 2223
 QY 587 -----LeuTyrAspArgGlnGlyLysAla 594
 Db 2224 TTTTTCACACTTGTGTGTTCCCTTTCAGCAACAATAATGGATGACATCTCTGGACCTATC 2283
 QY 595 IleLysAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro 614
 Db 2284 CAAATGCAACCAACATGAAGCAACACTTTCAGTTTGGGGCAGGACAGTCCAACCA 2343
 QY 615 GlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIleThrTyr 634
 Db 2344 AATTAGTGTGAATCCCGTCTCGTATACGATTTAGGCATCAAGGATTACCTCAACTTC 2403
 QY 635 LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSer 654
 Db 2404 TTATGCTCACTGGATCAACATCAAAATCTCTGTATTTCTCGGCAACAACCTTACT 2463
 QY 655 CysAsnGlyIleLeuArgLysAsnProGlyPheSer-----LeuAsnTyrProSer 671
 Db 2464 TGTTCA-----AGCCTTAAATCAGTCTTGTATACTCACTATCTCTTCC 2508
 QY 672 IleAlaVal-----IlePheLysArgGlyLysThrThrGluMetIleThrArgArgValThr 690
 Db 2509 ATCAGATTCCAAACTTAACATCAAGCAAACTCACC-----GTTCAAGAGCTGTAAA 2562
 QY 691 AsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGlyLys 710
 Db 2563 AACGTTGACGACCT---TCGATGTATACCTCAAGGTGAATAACCCACAGGGTGTAT 2619
 QY 711 ValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArg 730
 Db 2620 CTCGGGTGAAGCCACAGAGCTTGAATTTACCAAGGTGGAGAGCAAAAGCGTTTAAA 2679
 QY 731 ValTrpPheValLeuLysLysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGly 750
 Db 2680 GTGATCTTGTG-----AAAGTAAGGAATGTGGCTAAGGGCTATGTCTTTGGA 2730
 QY 751 GlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal 770
 Db 2731 GAATTTGTTTGGTCCAGACAGAAGCAT-----CGTGTGAGAAGTCCCATTTGTGTG 2781
 QY 771 ThrLeu 772

Db 2782 ARGCTC 2787
 RESULT 10
 US-09-938-842A-338
 ; Sequence 338, Application US/09938842A
 ; Publication No. US20040009476A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Kreps, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; TITLE OF INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SRIPI300-3
 ; CURRENT APPLICATION NUMBER: US/09/938,842A
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227,866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264,647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300,111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 338
 ; LENGTH: 2958
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-09-938-842A-338

Alignment Scores:
 Pred. No.: 1.29e-111 Length: 2958
 Score: 1074.00 Matches: 259
 Percent Similarity: 47.38% Conservatives: 102
 Best Local Similarity: 33.99% Mismatches: 204
 Query Match: 26.73% Indels: 198
 DB: 12 Gaps: 14

US-09-806-767-2 (1-775) x US-09-938-842A-338 (1-2958)

Qy 185 ArgLysLeuIleGlyAlaArgPheIleArgGlyHisArgValAla----- 200
 Db 572 AGGAACCTGATGAGGACGAGTACTCAACAGAGGCTCACCACCGCGGTAGACATCTC 631
 Qy 201 AsnSerProGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThr 220
 Db 632 AACTCAAGCTTCGACTCTCCA-----CGTGATCTCGAC 664
 Qy 221 GlyHisGlyThrHisThrAlaSerThrValGlyGlySerSerValSerMetAlaAsnVal 240
 Db 665 GGCCATGGCTCCCACTCTCTCCACCGCGCGCGAGATTGTTCTCGGAGTCAGCATC 724
 Qy 241 LeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyr 260
 Db 725 TTCGGCCCAAGCAACGCGACGCGCAAGGTGGCTCCCTCGCGCCAGAGTGGCGCGTAC 784
 Qy 261 LysValCysTrp-----PheAsnGlyCysTyrSerSerAspIleLeuAlaAla 276
 Db 785 AAAGTCTGTGCGCGCCAGCTCAAGGCAACGAATGCTACCGCAGACGCTCTCGTGCT 844
 Qy 277 IleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPhePro 296
 Db 845 TTCGACGCTGCATACACGCGGAGCTGACGTATATCCGTCTCTCTAGCGCGGAGCCT 904
 Qy 297 IleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIle 316
 Db 905 ACCTCGTTTAAACGACAGTGCCTGCTATTGTTCTTTCACGCGCTAAGAGCGGATC 964
 Qy 317 SerValIleCysAlaAlaGlyAsnGlyProIleGluSerSerValAlaAsnThrAla 336
 Db 965 GTGGTGTGTTCTCTGCGGAACTCTGGACCGGACAGATAGTACGGTTTCTAATGTCGG 1024
 Qy 337 ProTrpValSerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValValArg 356

Db 1025 CCATGGCAGATTACCGTTGGTGTCTAGCACCATGACCGTGAGTTTGTAGCAATCTTGT 1084
 Qy 357 LeuAlaIleGlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsn 376
 Db 1085 CTCGGTAATGGAACA-TTACAAGGTTAGATTAAGTCCAAGATATGATGATCATATCT 1143
 Qy 377 AlaGlyArgGluValGluVal-----IleTyrValThrGlyGly 389
 Db 1144 CTATCTCAATCTTGAANAATCTTGAATGAGCTTGAATCTTTTCATCTCGTAGG 1203
 Qy 390 AspLysGly----- 392
 Db 1204 GACAAGCTTCTCATCCAGCGCATTCACATGCTAAGTCTTATCCAAATTATGGCATCTG 1263
 Qy 392 ----- 392
 Db 1264 TAAATGCAAAAGCTAAGAACGCTTACGCTTTGGATGCGTGAGTTGGCTAATCTCTTAC 1323
 Qy 393 -----Ser 393
 Db 1324 TTAATTAGCCCAATGTTTATGTTAATGTTAATGTTAATGTTAATGTTAATGTTAATGTTA 1383
 Qy 394 GluPheCysLeuArgGlySerLeuProArgGluGluLeuArgGlyLysMetValIleCys 413
 Db 1384 CAATTATGCAAACTTGGATCGCTTACCCCTATAAAGACAAAGGGAAGATATTAGTCTGT 1443
 Qy 414 AspArgGlyValAsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyVal 433
 Db 1444 CTCAGAGGCGACAAACGGGAGGTCGAAAGGTCGGGCGCTTTAGGGGGTGGGATA 1503
 Qy 434 AlaMetIleLeuAlaAsnThrGluIleAsnGlnGluAspSerIleAspValHisLeu 453
 Db 1504 GGGATGGTCTTGAACAACACTTATGCACCGGGAATGATTTACTCGCTGACCCCATCTGTC 1563
 Qy 454 LeuProAlaThrLeuIleGlyTyrThrGluSerValLeuLysAlaTyrValAsnAla 473
 Db 1564 CTTCCAGCCACACAGCTCACTTCTAAGATAGTTTCTGTGTCAAGATATATCAGCCAA 1623
 Qy 474 ThrVal----- 475
 Db 1624 ACCAAGTAGGATTCTTGATTCACACTTTACCTAACTTTGTTCAAGTTCTCAAGAAAAAC 1683
 Qy 476 -----LysProLysAlaArgIleIlePheGlyGlyThrValIleGly 489
 Db 1684 TTTATTTTGTATGTAGGAGCCGATTGACACATTACTCTTCAAGACAGATTGGGA 1743
 Qy 490 ArgSerArgAlaProGluValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnPro 509
 Db 1744 CTAAACCCAGCGCTGTAAATGGCTTCAATTTCTTCAAGAGTCCAGCATAGTGGCTCCT 1803
 Qy 510 SerIleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaIleTrpGln 529
 Db 1804 CAGATTTCTGAGCTGACATAACTGCTCTGTGTGAGTGATGATCGCTGCTACACTGGA 1863
 Qy 530 AsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsnPheThrValMetSer 549
 Db 1864 GCAGTCTCTCAACAACAAACAGTTTGTATCTCTCGACGACTTCTGTTCATGCTATTTC 1923
 Qy 550 GlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTyr 569
 Db 1924 GGAACCTCTATCTCTGTCTCTCATATCTCTGCGCATTCGGGTCTTCTCAAAACGCTTAT 1983
 Qy 570 ProAsnTyrSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAsp----- 586
 Db 1984 CTTTCTTGAGCCCGCGAGCTATCGCTCTGCCATCATGACTACCGTAAGAATGTATT 2043
 Qy 586 ----- 586
 Db 2044 CGATTCTAGAGCTTACATTTGACAAATGGAAGTTTAATCTTATTTCACCTATA 2103
 Qy 586 ----- 586

Db 2104 AACACATTAACTACTGCTGCTGAGCTTATAACCAAAATCTGTCTTCATGGCCAT 2163
Qy 586 -----
Db 2164 GTTCTGTTTCTTAGTTAGTTGGAAATATTGGAATGAGTTCTTCTGAATCTTACGT 2223
Qy 587 -----
Db 2224 TTTTTCACACTTGTGTTTCCCTTTCAGCAACAATAATGATGACATTCCTGGACCTATC 2283
Qy 595 IleyAspGlyAsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisValAsnPro 614
Db 2284 CAAATGCAACCAACCAAGAGCAACACCTTTCAGTTTGGGGCAGACACGTCCAAACCA 2343
Qy 615 GlnLysAlaIleAsnProGlyLeuValTyrAsnLysGlnProValAspTyrIleThrTyr 634
Db 2344 AATTAGCTGTGAATCCCGTCTGTATACGATTAGGCATCAAGGATTACCTCAACTTC 2403
Qy 635 LeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSer 654
Db 2404 TTATGCTCAGTTGATACATGATCAACAAATCTCTGTATTCTCGGGCAACAACCTTTACT 2463
Qy 655 CysAsnGlyIleLeuArgLysAsnProGlyPheSer-----LeuAsnTyrProSer 671
Db 2464 TGTTCAC-----AGCCCTAAATAATCAGTCTTGTAACTCAACTATCTCTCC 2508
Qy 672 IleAlaVal-----IlePheLysArgGlyLysThrThrGluMetIleThrArgArgValThr 690
Db 2509 ATCAGAGTCCAAACTTAACATCAAGCAAACTCACC-----GTTTCAAGGACTGTTAAA 2562
Qy 691 AsnValGlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLys 710
Db 2563 AACGTTGGACGACCT---TCGATGTATACCTCAAGGTGATACCAACACAGGGTGTAT 2619
Qy 711 ValIleValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArg 730
Db 2620 GTCGGGTGAAGCAACAGAGCTTGAATTTTCCAAAGTTGGAGAGCAAAAGACGTTTAAA 2679
Qy 731 ValTrpPheValLeuLysLysLysAsnArgGlyLysValAlaSerPheAlaGlnGly 750
Db 2680 GTGATCCCTGTG-----AAAGTAAGGGAATGTGGCTAAGGCTATGTGTTGGA 2730
Qy 751 GlnLeuThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerVal 770
Db 2731 GAATGGTTTGTGTCAGACAAGAGCAT-----CGTGTGAGAAGTCCCATTTGTGCTG 2781
Qy 771 ThrLeu 772
Db 2782 AAGCTC 2787

RESULT 11

US-10-260-238-4334

; Sequence 4334, Application US/10260238

; Publication No. US20040016025A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, Paul R.

; APPLICANT: Moughamer, Todd G.

; APPLICANT: Briggs, Steven P.

; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane

; APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyaki

; APPLICANT: Kreps, Joel

; APPLICANT: Provart, Nicholas

; APPLICANT: Ricke, Darrell

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION

; FILE REFERENCE: 60111-NP

; CURRENT APPLICATION NUMBER: US/10/260,238

; CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/325,448

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 4334
; LENGTH: 1934
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-10-260-238-4334

Alignment Scores:
Pred. No.: 3,61e-97 Length: 1934
Score: 945.00 Matches: 244
Percent Similarity: 53.79% Conservative: 111
Best Local Similarity: 36.97% Mismatches: 221
Query Match: 23.52% Indels: 84
DB: 12 Gaps: 25

US-09-806-767-2 (1-775) x US-10-260-238-4334 (1-1934)

Qy 141 IleIleGlyValLeuAspThrGlyValTrpProGluSerProSerPheAspThrGly 160
Db 3 ATCATCGAGTTGTTGACACCGGATCTGGCCGGAGTCAAGAGCTTCAGCGATGTAGG 62
Qy 161 MetProSerIleProArgLysTrpLysGlyIleCysGlnGluGlyLysSerPheSerSer 180
Db 63 TAGGACCGTCCACGAGGTGGAAAGGGAGTGCACAGCGCGCGGACATGGGGCATC 122
Qy 181 SerSerCysAsnArgLysLeuIleGlyAlaArgPhePheIleArgGlyHisArgValAla 200
Db 123 AACAACTGCAGCGGAAATCATCGCGCGGTTCTAC-----ACCGCT 167
Qy 201 AsnSerProGluGluSerProAsnMetProArgGluTyrIleSerAlaArgAspSerThr 220
Db 168 GGGGTCCCGACGAG-----TACCTCAAAGGAGTCTGCTGCACCGCGGACCATCTAC 221
Qy 221 GlyHisGlyThrHisThrAlaSerThrValGlyLysSerSerValSerMet----- 237
Db 222 AGCCACGCGCACCATTTGCTTCCACCGAGTGTGCTCAGCGTAGGCTGGACCGCGGAG 281
Qy 238 ---AlaAsnValLeuGlyAsnGlyAlaGlyValAlaArgGlyMetAlaProGlyAlaHis 256
Db 282 GAGGCTAGTTCACCGCTTCTCGGGGGTGGCGGGGAGCGCACCGCGCGCTCGC 341
Qy 257 IleAlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAspIleLeuAla 276
Db 342 ATCGGGGTACAGTCCATCTTGTCTGCTACCGTACCGTGTATCAGCCTGTGTGCGCGC 401
Qy 277 IleAspValAlaIleGlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPhePro 296
Db 402 ATTGATGACCAATCCACGCGGTGGATGTATTGTGCTGTCCCTC-----ACT 452
Qy 297 IleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIle 316
Db 453 GTGCCG---TAGAGAAC-----TCGTACGGTGCCTGCACCGCGTCCAGAGGGGATC 503
Qy 317 SerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAla 336
Db 504 ACTGTGTATCTCCGGGGAACACGAGGACCCAGCGCGCAGACCGTTCGAGGATACCTCG 563
Qy 337 ProTrpValSerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValValArg 356
Db 564 CCTTGGGTCTCATGTAGCGCGGAGCAAGATCGACCGTCTCTCCCGACGTGTATCAGC 623
Qy 357 LeuAlaAsnGlyLysLeuLeuTyrGlyLysLeuTyrProGlyLysGlyIleLysAsn 376
Db 624 TTGGGAACAACAGCAAAATGTGGCCAGTCTCTCTAT-----TACCAAGTGAAGAAC 677
Qy 377 AlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGlySerGluPheCys 396
Db 678 ATCAACTCATCATCAC-----ACCGTAGAAATTTACAAAGCTTATATGT 725
Qy 397 LeuArgGlySerLeuProArgGlu-----IleArgGlyLysMet 410

```

Db      726 GCACATCTGTGCGGACATGAGGTTCTTCATCATATTGCACTGAAAGGGGTATC 785
Qy      411 ValIleCys-----AspArgGlyValAsnGlyArgSerGluLysGlyGluAlaVal 427
Db      786 CTGGTGTCTTAATGATATGATGATGCGGATCAATCCATCATGATGATTTCACCATGCGGG 845
Qy      428 Lys-----GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGln 444
Db      846 AAGTACGTGTGGAAGCGGGGATCGGGCTCATCTTCGCG-----CAATACACAACG 899
Qy      445 GluGluAspSerIleAspValHisLeuLeuProAlaThrLeuIleGlyThrGluSer 464
Db      900 GACACCTGAGTCCACAGTGGCTTACTGCCGGG-----ATGCGCTGC 944
Qy      465 ValLeuLeuLysAlaThrValAsnAlaThrVal-----Lys 476
Db      945 GTTATCGTCGACATCTACATCTCCAGGAGGATAGCCATACGATTCGACTCAAGCTCG 1004
Qy      477 ProLysAlaArgIleIlePheGlyGlyThrValIleGlyArgSer---ArgAlaProGlu 495
Db      1005 CCCACACAAAATCGAACCCGCGCCGTCATGAGCAAGATGTGCTGCTTCGAA 1064
Qy      496 ValAlaGlnPheSerAlaArgGlyProSerLeuAlaAsnProSerIleLeuLysProAsp 515
Db      1065 GTGGCGTCTCTCTTCGAGAGGCCATCGCTGATTACCGCGCTTCATCAGCTGAC 1124
Qy      516 MetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsnLeuGlyProThrGly 535
Db      1125 GTAGCTCGCGCGGAGCCACATCTTAGCAGCA-----GTGGGAGAT----- 1166
Qy      536 LeuProTyrAspSerArgArgValAsnPheThrValMetSerGlyThrSerMetSerCys 555
Db      1167 -----TCTTACGTAACTAAGTCAGAACGCTCCATCGGACGC 1202
Qy      556 ProHisValSerGlyIleThrAlaLeuIleArgSerAlaThrProAsnTrpSerProAla 575
Db      1203 CCGCATCGCGGGTGTGCGCGCGCTGCTCAAGCTCTGCACCGGATGGTCTCTGCT 1262
Qy      576 AlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArgGlnGlyLysAlaIle 595
Db      1263 GCATAAAGTCAGCCATCGTCACCATCGCATGTACTGATGACACGCGATCGCGATC 1322
Qy      596 ---LysAspGly-----AsnLysProAlaGlyValPheAlaIleGlyAlaGlyHisVal 612
Db      1323 CTAGCGGAAGGACTACCGCGGAGTGGCGGATCCATTCGACTATGAGCGCGGAACATC 1382
Qy      613 AsnProGlnLysAlaIleAsnProGlyLeuValTyrAsnIleGlnProValAspTyrIle 632
Db      1383 AACCCATCGCGGCGAGCTCATCCGGGCTAATTATGACATTCATCCACACGATTAACAAC 1442
Qy      633 ThrTyrLeuCysThrLeuGlyPheThrArgSerAspIleLeuAlaIleThrHisLysAsn 652
Db      1443 AAATTCCTT-----CAATCGGCAATGTCAAGAGGGGATATCTCGGATC----- 1487
Qy      653 ValSerCysAsnGlyIleLeuArgLysAsnProGlyPheSerLeuAsnTyrProSerIle 672
Db      1488 GCAGCTGCAAC-----ACACCTCTATGCCAGAGTATAACCTCATCTGCGCTCATC 1541
Qy      673 AlaValIlePheLysArgGlyLysThrThrGluMetIleThrArgArgValThrAsnVal 692
Db      1542 TCGGTTCGGATTTTAGAGCGCGATCACC-----GTCTCGAGGACGGTAACCAATGTC 1595
Qy      693 GlySerProAsnSerIleTyrSerValAsnValLysAlaProGluGlyIleLysValIle 712
Db      1596 GCGAGCTCCATCTGTGTACACCGCTGTAGTCCAGAGCCCGACTGGAGTCAATATGGAG 1655
Qy      713 ValAsnProLysArgLeuValPheLysHisValAspGlnThrLeuSerTyrArgValTrp 732
Db      1656 GTTTTCCGGATGTCTGGTGTCTGATGCTCGGACAAAGTCCGGAAATTCGAGGTGAAG 1715
Qy      733 PheValLeuLysLysLysAsnArgGlyGlyLysValAlaSerPheAlaGlnGlyLeu 752

```

```

Db      1716 CTGTCACTCTCTCGAACTTTCAAGGGGGC-----TACACGTTTGCAGCATC 1763
Qy      753 ThrTrpValAsnSerHisAsnLeuMetGlnArgValArgSerProIleSerValThrLeu 772
Db      1764 ACTTGGCAGATGACCGA-----CAAGTAGTAGGATTCCAGTTGCGACCCGATT 1814

RESULT 12
US-09-754-853A-4/c
; Sequence 4, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 4
; LENGTH: 513509
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (111805)..(113968), (114684)..(115204)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(513509)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 318013_region_A3
US-09-754-853A-4

Alignment Scores:
Pred. No.: 5,25e-89 Length: 513509
Score: 909.00 Matches: 285
Percent Similarity: 29.98% Conservative: 108
Best Local Similarity: 21.74% Mismatches: 216
Query Match: 22.62% Indels: 702
DB: 11 Gaps: 22

US-09-806-767-2 (1-775) x US-09-754-853A-4 (1-513509)
Qy      145 LeuAspThrGlyValTrpProGluSerProSerPheAspAspThrGlyMetProSerIle 164
Db      195773 TTAAGCTAGGTGTTTGGCCCTGAATCAAGAGCTTTAGTGAGCAAGGTTTGGGACCAATT 195714
Qy      165 ProArgLysTrpLysGlyIleCysGlnGluGly----- 175
Db      195713 CCATCAAGTGGAGAGGAAATCTGTGATAACGGAATTGATCATATCTTTTCACGCAACAGG 195654
Qy      175 ----- 175
Db      195653 TACCATCTTCCTTAGAATGGAATTAACACTATTAACTATTTGTGAGTCATTATAATGGA 195594
Qy      175 ----- 175
Db      195593 TATAATAAAAATGTTTCTTATCTAATTGTTTAACTTCTGTAGTCATATGCCTTTAAATTG 195534
Qy      175 ----- 175
Db      195533 GCACCTCTTTATAGGTGAATACTACATGAAGGGTTAAGCACTCTTTGAACGACAGACA 195474
Qy      175 ----- 175
Db      195473 TGTTTCTTTTCGTAFTCTTATCTTTATAGTACTAGTGTGTTTTCACACCAACGACCAAT 195414
Qy      176 -----GlusPheSerSerSerSerCysAsnArg 185

```

```
Db 195413 CTTTCATGATTTTGTGTTTGAATAATTAAGAAACGTTAAGTGGAGATGTTGAATGCAGG 195354
Qy 186 LysLeuIleGlyAlaArgPheIleArgGlyHisArgValAlaAenSerProGluGlu 205
Db 195353 AAGCTTAATAGGGGAAGGACTTCAACAAGAGCTACGCTCAGTGGCGGGCCCACTGAAC 195294
Qy 206 SerProAenMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHis 225
Db 195293 TCC-----TCCTTTGACTCGCGCCAGCTGACACAGGAGGCCACGGAGCCAC 195249
Qy 226 ThrAlaSerThrValGlyGlySerSerValSerMetAlaAenValLeuGlyAsnGlyAla 245
Db 195248 ACCCTATCCACAGCTGGCGGGAACAATGTCACAGTGTACGCTTCAGCGTCTTCGGCCAGGCCAG 195189
Qy 246 GlyValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrp--- 264
Db 195188 GGAACAGCCAAAGGTGGCTACCAATGCGACGTGTGCGAGCTACAAGGTCTGCTGSCCC 195129
Qy 265 -----PheAsnGlyCysTyrSerSerAspIleLeuAlaIleAspValAlaIle 281
Db 195128 CCCCTCGAGGCGAGGAGTGTCTCGACGCTGATATCTTGGCGGCTTCGACCTCGCCATC 195069
Qy 282 GlnAspLysValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAsp 301
Db 195068 CACGACGCGTCGATGTCCTCTCGCTCTCGCTCGGCTCTTCTCCACGTTTTTCAAG 195009
Qy 302 AspThrIleAlaIleGlyThrPheArgAlaMetGluArgGlyIleSerValIleCysAla 321
Db 195008 GATAGTGTTCGATCGGATCTCTTCATCGCGCCAGGCTGGCGTGTGGTGTGTCTCT 194949
Qy 322 AlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTyrValSerThr 341
Db 194948 GCTGCAACAGTGGCGCGGGAAGCCACTCGGGAATACTCGCACCTCGCATGTTTACG 194889
Qy 342 IleGlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAlaAsnGly--- 360
Db 194888 GTTGTGCGACACCATGATGATCGACAGTTCGCCACTTACGTTGTCTTGGTAACGATATT 194829
Qy 361 -----LysLeu 362
Db 194828 ACCTTCAAGGTTGTGTCTCTTCCACACACATGATGCAAAATTATGACTAAACTT 194769
Qy 362 ----- 362
Db 194768 TACGATAAATACTCAATTCATCAAGTGTGTTGAATAATAATTAATAGTTCGATGATA 194709
Qy 362 ----- 362
Db 194708 TTCAAATTTAATCACTCAATAGGTACAAAGTCGACTAATTTGTCATATTTTTTACAGA 194649
Qy 363 -----LeuTyrGlyGluSerLeuTyrProGly 371
Db 194648 TTAAATTTGTATGCCCACTAATTTATCATATTTTTTACGGATTAATTTGTATTATCA 194589
Qy 372 Lys----- 372
Db 194588 TCTTAGAGACTAATTTATCACCGCATCATCTTTGGAGGATAAATTAATTAATTAACAAA 194529
Qy 372 ----- 372
Db 194528 CTTTGTATTAGCTCTTAGTTGCTGTAGTGTCTTACATTTATTTGTTGACTAAATTTTG 194469
Qy 373 -----GlyIle--- 374
Db 194468 AGTTTATTTCGGTAAATTTATGTAATTTTTTAAATGTAACCTTTTAAATTAATTCGGGTGAAA 194409
Qy 374 ----- 374
Db 194408 TTCTAATTCGATTAGGAAACATTAAACAATAGTAGTACGAAATTAACCTAAGATTTTGTAG 194349
Qy 374 ----- 374
Db 194348 TATAAATTTATGTAGACTTTTGAGTATAGGGACTATAAATTTATGTGAGTCTAAGAGAGAA 194289
```

```
Qy 374 ----- 374
Db 194288 AGATATACTTTTTTAAAAAATTTAATAATAGCTTCTGCAACTAGTTAGGATTTGTCATTT 194229
Qy 374 ----- 374
Db 194228 TTTTAATTAGTTGAAAAGATTCATGAATTTAAATTTTGATATTTTTTTCTTTTGTGTTTCAACAT 194169
Qy 374 ----- 374
Db 194168 AACCATGTTATTTAATAACAGGAAAGTACGTGAAAGTCATGATTAAAGTGTGATTGTAG 194109
Qy 375 -----LysAsnAlaGlyArgGluValGluVal 383
Db 194108 TTGCAGTTGTGTGTGTGTGTACGAAATTTCTGAAAAAATAATTGGACAAATGTACGAAATTT 194049
Qy 384 IleTyrValThrGly-----GlyAsp--- 390
Db 194048 GATTTCCAAATACAGATGTGGCCAGTCGAGGATTTTGCATTTTATAAACACTATACAGTGA 193989
Qy 391 -----LysGlySerGluPhe--- 395
Db 193988 CACCATAATCTCGGTGCTTGATTGTAAGGGGAAAGCTTATCAGCTACAAAATTTGGCAC 193929
Qy 395 ----- 395
Db 193928 ACAAGTTCTATCCAAATTATTAAAGCCACAGATGCTAAATTTGCGAGTGCAGAGCTGAAG 193869
Qy 395 ----- 395
Db 193868 ACGCGTAAGTATCAAAAAATATGAAACTCGTTATAGTCATCATGTTGTTGAATGATTGACAT 193809
Qy 396 -----CysLeu 397
Db 193808 AATTAACATTCGACTATATATGTTAAATAAATTTATCTCAAGCAGGGTGTGTGCCAG 193749
Qy 398 AsgGlySerLeuProArgGluGluIleArgGlyLysMetValIleCysAspArgGlyVal 417
Db 193748 AATGGACCTCGATGCCCAACAGCGGAGGAAAGATCGTGGTGTGTCTCGAGGAATA 193689
Qy 418 AsnGlyArgSerGluLysGlyGluAlaValLysGluAlaGlyValAlaMetIleLeu 437
Db 193688 AATGCAAGAGTGCACACAGGAGCAAGCTTTTCTAGCTGTGTGCTGCGGAATGTCTCT 193629
Qy 438 AlaAsnThrGluIleAsnGlnGluGluAspSerIleAspValHisLeuLeuProAlaThr 457
Db 193628 GCTAATGCAAGACTACTGGGAATTAAGCGGATTCCTCATGTCTCTCTCTCTCTCT 193569
Qy 458 LeuIleGlyTyrThrGluSerValLeuLeuLysAlaTyrValAsnAlaThr--- 474
Db 193568 CATATCAATTTTACTGATGGTAGTGTCTGCTTTAATTACATCACTCAACCAAGTAAGAT 193509
Qy 474 ----- 474
Db 193508 ATCTACCACCTAGTCTGCCACATATTAAATAATAATTAATTAATTAATTAATTAATTAAT 193449
Qy 475 -----ValLysProLysAlaArgIleIlePheGlyGlyThrVal--- 487
Db 193448 GCCTACTTTGGGTTTGGCACTATATACTATATGACATACCACTTATATTTTGGCAATCAGAA 193389
Qy 488 -----IleGly 489
Db 193388 TATATGGATTTGCTATTAACCTTATAATTTGTGATAGTTGGTATTTTTTTTGTAGCAATACAA 193329
Qy 490 ArgSerArg--- 492
Db 193328 AGATCCCAATGATAATCAATCAATCTCAGTTTTTCAACTTGAACCTTGACTTACTTTAAAT 193269
Qy 492 ----- 492
Db 193268 TTATGTATTGTAGGTTCCAGTGGCATATATTACACATCCAAAGACTCAATTTGGATACT 193209
```

```
Qy 493 -----AlaProGluValAlaGlnPheSerAlaArgProSerLeuAlaAsnProSer 510
Db 193208 AAGCCAGCCCACTTATGGCAGCATTTTCATCAAAAGGACCAATACTATGTGGCCAGAG 193149
Qy 511 ILeuLys-----513
Db 193148 ATCCCTAAAGGTGTGATTACTTACAGTGTATTATGGCTTGTAAATTTGATGTAATGTGTAG 193089
Qy 514 -----ProAspVetIleAlaProGlyVal 521
Db 193088 ATATCTAAGAACTCTTAATTTGTCCAAATATATTCAGCCTGATATCATCTGCACAGCGGTG 193029
Qy 522 AsnIleAlaAlaLysProGlnAsnLeuGlyProThrGlyLeuProTyTrAspSerArg 541
Db 193028 TCAGTTATAGAGCCTATATCTAGAGCCCAAGGACCAACCAATCAAGTGTTCGACAGCCG 192969
Qy 542 ArgValAsnPheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIle 561
Db 192968 AGGATTCGGTTTAATCTAGCTCAGGCACATCAATGTCTGATGCCCTCATGTTTCAGGCATT 192909
Qy 562 ThrAlaLeuIleArgSerAlaTyProAsnTrpSerProAlaAlaIleLysSerAlaLeu 581
Db 192908 GTGGGACTGTAGAGCCTTATATCTCTACGTGGAGTACCGCTGCAATTAATCAGCAATC 192849
Qy 582 MetThrAlaAsp-----586
Db 192848 ATGACCACAGGTGAGTGAGTGTTCTTCATGGTTTCTCATGCTTCAAACTCTTTT 192789
Qy 586 -----586
Db 192788 TCAAGCCAACTACAATAAATGTCCAGAAATATAAAGTTAGATATATAGAAAATA 192729
Qy 586 -----586
Db 192728 ACTTAGATATAGTTTCTTAATCTGCTAGTGTATAAAGGTTTTCATTAAAGATTCT 192669
Qy 586 -----586
Db 192668 GCAAGTGCAAACTCTAATCTTAATTTTAAAAAAGGATTTACAGTACTTAAAGAACT 192609
Qy 587 -----LeuTyAsnArgGlnGlyAlaIleLysAspGly 598
Db 192608 ATATGTAAATGTAAAGTATGTCTTAAAGATTTATGAGGGAGGTGAACCTCCAATTCT 192549
Qy 599 Asn-----LysPro-----601
Db 192548 AATTGGAGACGACCACTAGATATCTAAGGAATGTATCTAAGTTTGTCCATACATAT 192489
Qy 599 -----599
Db 192488 ATAGAAAGCAATTTGATTAGTGAGTAAAGTAGCTATCAAAAATGTATGTTTGTCTT 192429
Qy 600 -----LysPro-----601
Db 192428 CTGCTACTTTCAGTACACACTAGACAAATGAGTGGAGCCACTACTGAATGTCTGTAT 192369
Qy 602 -----AlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsn 619
Db 192368 GCGAAGGCGACACCATTCAGTACGGGCGAGGCGATGTTCAACCAACAGAGCAATGGAT 192309
Qy 620 ProGlyLeuValTyAsnIleGlnProValAspTyIleThrTyLeuCysThrLeuGly 639
Db 192308 CCGGGCTAGTTTATGACATAACAAATGTATGATTTACCTCAACTTCTTATGTCTTAGGA 192249
Qy 640 PheThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeu 659
Db 192248 TACACGACGACACAGATCTCAGTGTTCACAGAGTCCATATAAATGCCGC-----192198
Qy 660 ArgLysAsnProGlyPheSerLeuAsnTyProSerIleAlaValIlePheLysArgGly 679
Db 192197 AAGAAATTTAGTCTCTCAACCTGAACATATCCCTCAATCACTCAGTCCCAAACTCTCTGGG 192138
Qy 680 LysThrThrGluMetIleThrArgValThrAsnValGlySerProAsnSerIleTy 699
```

```
Db 192137 TCAGTGCACA-----GTTACAGGACATTTGAAAAAATGTTGGTTCTCCAGGACT---TAC 192087
Qy 700 SerValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuVal 719
Db 192086 ATTGCTCATGTTCAAAACCCCGTATGGAATCACCGTTTCTGTGAAGCCACGATCTTGAAG 192027
Qy 720 PheLysHisValAspGlnThrLeuSerTyArgValTrpPheValLeuLysLysAsn 739
Db 192026 TTCAAAATGTTGTGAAGAGAGAGAGCTTTAAATTAACCTTC-----AAGGCCATG 191976
Qy 740 ArgGlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsn 759
Db 191975 CAAGAAAGGCCCAATAATACTACGCGTTTGGAAAACTGATATGTCAGATGGCAAGCAT 191916
Qy 760 LeuMetGlnArgValArgSerProIleSerVal 770
Db 191915 TAC-----GTCAAGAGTCCAATAGTTGTG 191892

RESULT 13
US-09-938-842A-64
; Sequence 64, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kresps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: SPRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 64
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-64

Alignment Scores:
Pred. No.: 5, 8e-92 Length: 2451
Score: 901.00 Matches: 269
Percent Similarity: 48.25% Conservative: 130
Best Local Similarity: 32.53% Mismatches: 327
Query Match: 22.42% Indels: 101
DB: 10 Gaps: 28

US-09-806-767-2 (1-775) x US-09-938-842A-64 (1-2451)
Qy 6 PhePheLeuCysIle-----IlePheLeuLeuPheCysSerSer 19
Db 28 TTTTTCATGTTTCTTCTACTGTGCACAGAGATTACATGTGACTATGGAAGGAGAA 87
Qy 20 SerGluIleLeuGlnLys-----GlnThrTyIleValGlnLeuHis 33
Db 88 CCAATCATTTAGTTACAAAGTGGTGATAATGGAATTTGAAGCAACTGCTGTGAATCTCAT 147
Qy 34 ProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrpHisLeuSerPheLeu 53
Db 148 GAGAAAAATGATGATACAACTGACAACTGATGATCATCATCGCCCTTCACCTTTGAGAGAG 207
Qy 54 GlnGluAlaValLeuGlyValGluGluGluGluGluGluGluGluGluGluGluGluGlu 73
Db 208 CATGATATGCTTCTTGAATG-----CTCTTTGTGGAGGATCTTACAAAAGCTTTTAC 261
Qy 74 SerTyGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluIle 93
```

Db 262 AGCTATAAACACCTCATTAATAGTTTTCAGCTCATGTATCCCTCATCAGCGGAAATG 321
Qy 94 LeuArgTyrSerProGluValAlaValArgProAspHisValLeuGlnValGlnThr 113
Db 322 CTTCGTCCGGCCCTGTGTGAATCTCTGTAGATAGAGATTGGAAGATTAGGAACCTACT 381
Qy 114 ThrTyrSerTyrIysPheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrp----- 131
Db 382 ACACATACACCAAGTTTGGGATTACCA-----ACTGATGTTTGGCCCACT 429
Qy 132 -----SerIysSerArgPheGlyGlnGlyThrIleIleGlyValLeuAspThrGlyVal 149
Db 430 GGTGGTGGTTATGATAGACGCGGGAAGATATTGTTATGGCTTTATTGACTCAGGGATA 489
Qy 150 TrpProGluSerProSerPhe-----AspAspThrClyMetProSerIleProArg--- 166
Db 490 TTTCACATACCAAGTTTGGCTCTCCACACACACAGTATGATGGCCCTCATCCT 549
Qy 167 LysTrpIysGlyIleCysGlnGlyGluSerPheSerSerSerCysAsnArgLys 186
Db 550 AGTTACAAAGGAATGTGAAGAGATCTCATACTAAGATCAGTTTTCATGATGGGAAG 609
Qy 187 LeuIleGlyAlaArgPheIleArgGlyHisArgValAlaAsnSerProGluGluSer 206
Db 610 ATCATAGGCGCAGCATTTGCTGTAAGCCGCTAAAGCAGCTGGTGCT-----TTCAAT 663
Qy 207 ProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThr 226
Db 664 CCGGATATC-----GACTTGTCTCACCATGGATGGATGGATGGATGGATGGATGGAT 717
Qy 227 AlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGly 246
Db 718 GCAGCTATTGTCAGCTGGGAACAATGGTATTCAGTGAGGATGCAGGTTATGAGTTTGA 777
Qy 247 ValAlaArgGlyMetAlaProGlyValAlaHisIleAlaValTyrIysValCysTrp----- 284
Db 778 AAGCAAGCGGATGGCTCTCTCGTGAAGGATGCTGTTTACAAAGCTCTACCGACT 837
Qy 265 PheAsnGlyCysTyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLys 284
Db 838 TTTGGAGC---TTTGTAGCTGATGTGTGGCTGCCATTGATCAGGCTGTTATGATGA 894
Qy 285 ValAspValLeuSerLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIle 304
Db 895 GTAGATATTGAGCTCTCGTGGTGGTCCAAACAGTCTCCAGCTACTACAAAGACAACA 954
Qy 305 AlaIleGlyThrPhe-----ArgAlaMetGluArgGlyIleSerValIle 319
Db 955 TTCTTGATCCATTGTGCTACACTTCTTGGGCTGTAAAGCTGGTGTGTTTGTGCT 1014
Qy 320 CysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpVal 339
Db 1015 CAAGCTGCTGAAACGGGTTCCCTTTCCGAAACTCTGTTTCATACAGCCCTTGATA 1074
Qy 340 SerThrIleGlyAlaGlyThrLeuAspArgPheProAlaValValArgLeuAlaAsn 359
Db 1075 ACTACTGTGGCTGTGCAATGTATGATCGCAGATACAAAAATCATCTGACTCTTGGAAAT 1134
Qy 360 GlyIysLeuLeuTyrGlyGluSerLeuTyrProGlyIysGlyIleIysAsnAlaGlyArg 379
Db 1135 GGTAAATAGTCTGCGGAATAGGATATTCTCTTACTCGACCTCATCTGTTTCATACAAG 1194
Qy 380 GluValGluAlaIleTyrValThrGlyGlyAspLysGlySerGluPhe----- 395
Db 1195 ATGGTTTCTGCAATGATGTTCTGCTTGTCTGTTCTGTTATGAAATCAATCCGTCGGAC 1254
Qy 396 CysLeuArgGlySer---LeuProArgGluGluIleArgGlyIysMetValIleCysAsp 414
Db 1255 TGCCAGGAAGCCAGAAAGTCTTGAACAAAAATTAGTCGAAGAAACATCTGCTTTTGGGA 1314
Qy 415 ArgGlyValAsn-----GlyArgSerGluIysGlyGluAlaValLys 428

Db 1315 TATTCTTTCAATTTGCTGTTGCTGCTTCCATCAAGAAAAGTTGCTGAACCTGCCAAG 1374
Qy 429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsn-GlnGluGluAspSe 448
Db 1375 CATCTAGCGCTGCTGTTTCTGTTCTGTTGCTGAAACGTTTCTCCAGGAACAAATTC 1434
Qy 448 IleAspValHisIleu---LeuProAlaThrLeuIle---GlyTyrThrGluSerValLeu 466
Db 1435 GATCCGCTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1494
Qy 467 LeuIysAlaTyrValAsnAlaThrVal----- 475
Db 1495 TTGATGATTAATCAATGTCAACAGTCAAGAGATTGGATGGGAAGGTTAAAGACTTT 1554
Qy 476 LysProLysAlaArgIlePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
Db 1555 AAAGCTGAAGAGCATCGGAGAGGTTTGAACCCATCTTCAAAATCCGACCTGAA 1614
Qy 496 ValAlaGlnPheSerAlaArgGlyPro-----SerLeuAlaAsnProSer 510
Db 1615 GTTGTCTGTTCTCAGCTCGAGGACCCCAATACCAAGATTTTCAGCTTTCAGGATGCTGAT 1674
Qy 511 IleLeuLysProAspMetIleAlaProGlyValAsnIleIleAlaAlaTrpProGlnAsn 530
Db 1675 CTCTCCTCAACCCAGATATTCTTCTCAGGCTCTTTAATATGCTGCTGCTGCTGCTGCTGCT 1734
Qy 531 LeuGlyProThrGlyLeuProTyrAspSerArgValAlaAsnPheThrValMetSerGly 550
Db 1735 ---GGAACAGACGAGGCTAATATATCGCGAA-----GGATTGCTCAATTTCTGCG 1785
Qy 551 ThrSerMetSerCysProHisValSerGlyIleThrAlaLeuIleArgSerAlaTrpPro 570
Db 1786 ACAGCATGGCTGCACACACATTCGCGGTATTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1845
Qy 571 AsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyrAspArg 590
Db 1846 CAATGGAGTCCAGCTGCCATTAATCAGCTTGTGACACTTCAACAGTCAATAGATAGA 1905
Qy 591 GlnGlyLysAlaIleLysAspGlyAsn-----LysPro 601
Db 1906 GCAGGAAGACCTCTCCAAAGACCAATATTCTGAACAGACAGACAGTAACTGTTTAAA 1965
Qy 602 AlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGly 621
Db 1966 GCAACTCGTTGATTATGGAAGGCTCATGTCATCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 2025
Qy 622 LeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeuCysThrLeu---GlyPhe 640
Db 2026 CTCATCTTGTATGTCAGGTTATGAGGATATATAGGTTCTTGTGCACACACCTGTTATC 2085
Qy 641 ThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArg 660
Db 2086 GATGCTCAGAGATAAAAACTTCA-----AACACTCCATGCAATTTCAAATGTT 2139
Qy 661 LysAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLys 680
Db 2140 CATCTCTCAAACTTC-----AACAGCCATCCATAGCCATC-----TCTCATCTGCTC 2187
Qy 681 ThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyrSer 700
Db 2188 AGAACACAAACCGTAACCAAGAGTACGATGTTGCGGAAGAAAGAAAGAACATACACA 2247
Qy 701 ValAsnValLysAlaProGluGlyIleValLysValIleValAsnProLysArgLeuValPhe 720
Db 2248 ATCATCATCGAGGATGGACCCCAATCCCATCGAAGTACCTCTCTGCTGCTGCTGCTGCTGCTGCT 2307
Qy 721 Lys---HisValAspGlnThrLeuSer----- 728
Db 2308 AGAGCGGGCGCTTCTAGAACCTTTTCGGTGTACTTAACAGTCAATCAGTCACTGCTGCTGCTGCTGCT 2367
Qy 729 TyrArgValTrpPheValLeuLysLysAsnArgGlyGlyLysValAla-----Ser 746
Db 2368 TATAGCTTCGGACAGGTTACATTGAAGGAAGCCGAGGCGCATAAAGTACTCTCTCCTGCTG 2427

Qy 747 PheAlaGlnGlyGln 751
 Db 2428 GTTGCTATGGACAA 2442

RESULT 14
 US-09-938-842A-64
 ; Sequence 64, Application US/09938842A
 ; Publication No. US20040009476A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Kieps, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; TITLE OF INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SCRIPI300-3
 ; CURRENT APPLICATION NUMBER: US/09/938,842A
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227,866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264,647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300,111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 64
 ; LENGTH: 2451
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; US-09-938-842A-64

Alignment Scores:
 Pred. No.: 5,8e-92 Length: 2451
 Score: 201.00 Matches: 269
 Percent Similarity: 48.25% Conservative: 130
 Best Local Similarity: 32.53% Mismatches: 327
 Query Match: 22.42% Indels: 101
 DB: 12 Gaps: 28

US-09-806-767-2 (1-775) x US-09-938-842A-64 (1-2451)

Qy 6 PhePheLeuCysIle-----IlePheLeuLeuPheCysSerSer 19
 Db 28 TTTTTCACATGTTTCTCCTCACTGTGACAGCAGAGATTTTACATTTGCTGACTATGGAGAGAA 87
 Qy 20 SerGluLeuGlnLys-----GlnThrTyrIleValGlnLeuHis 33
 Db 88 CCAATCATTAGTTACAAAGGTGTGATATGGAATTTGAAGCAACTGCTGTGGAACTGAT 147
 Qy 34 ProAsnSerGluThrAlaLysThrPheAlaSerLysPheAspTrpHisLeuSerPheLeu 53
 Db 148 GAGAAATTTGATCTACAGTGAACCTGCTGACATCATACCCCGTCACTTTGAGAGGAAG 207
 Qy 54 GlnGluAlaValLeuGlyValGluGluGluGluGluProSerSerArgLeuLeuTyr 73
 Db 208 CATGATATGCTTCTTGAATG-----CTCTTTGTGGAAGGATCTTACAAAAGCTTTAC 261
 Qy 74 SerTyrGlySerAlaIleGluGlyPheAlaAlaGlnLeuThrGluSerGluAlaGluIle 93
 Db 262 AGCTATAAACACCTCATTAATGTTTTCAGCTCATGTATCCCTCATCAGCGGGAATG 321
 Qy 94 LeuArgTyrSerProGluValValAlaValArgProAspHisValLeuGlnValGlnThr 113
 Db 322 CTTTCGTCGCGGCTGTGTGAAATCTGTGATAGAGATTGGAAGTTAGGAACCTTACT 381
 Qy 114 ThrTyrSerTyrLysPheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTyr----- 131
 Db 382 ACACATACACACAGTTTGGGATTACCA-----ACTGATGTTTGGCAACT 429
 Qy 132 -----SerLysSerArgPheGlyGlnGlyThrIleLeuGlyValLeuAspThrGlyVal 149
 Db 430 GGTGGTGGTTATGATAGACAGCGGGAAGATATTGTTGCTTTATTGACTCAGGGATA 489

Qy 150 TrpProGluSerProSerPhe-----AspAspThrGlyMetProSerIleProArg--- 166
 Db 490 TTTCCACATCACCACAGTTTGGCTCTCACCACACACACAGTAGTACTTATGGCCCTCATCCT 549
 Qy 167 LysTrpLysGlyIleCysGlnGluGlyGluSerPheSerSerSerSerCysAsnArgLys 186
 Db 550 AGTTACAAAGGGAATGTGAAGAGATCCTCATACATAGATCAGTTTCTGCAATGGGAAG 609
 Qy 187 LeuIleGlyAlaArgPhePheIleArgGlyHisArgValAlaAsnSerProGluGluSer 206
 Db 610 ATCATAGAGCGCAGCATTTTGTGAAGCGCTAAAGCAGCTGTGTCT-----TTCAAT 663
 Qy 207 ProAsnMetProArgGluTyrIleSerAlaArgAspSerThrGlyHisGlyThrHisThr 226
 Db 664 CCGATATC-----GACTTTGCTTCCCAATGATGCGATGACATGGAAGTACACACA 717
 Qy 227 AlaSerThrValGlyGlySerSerValSerMetAlaAsnValLeuGlyAsnGlyAlaGly 246
 Db 718 GCAGCTATTGCAGCTGGGAACATGGTATTCAGTGAGGATGCACGGTTATGATTGGGA 777
 Qy 247 ValAlaArgGlyMetAlaProGlyAlaHisIleAlaValTyrLysValCysTrp----- 264
 Db 778 AAGCAAGCGGATGGCTCTCGTCAAGGATTGCTGTTTACAAAGCTCTCTACCGACTT 837
 Qy 265 PheAsnGlyCysTyrSerSerAspIleLeuAlaIleAspValAlaIleGlnAspLys 284
 Db 838 TTTGAGGC---TTTGTAGCTGATGTGTGCTGCCATTGATCAGGCTGTTTCATGATGGA 894
 Qy 285 ValAspValLeuSerLeuGlyGlyPheProIleProLeuTyrAspAspThrIle 304
 Db 895 GTAGATATTTTGAAGTCTCTGGTTGGTCCAAACAGTCTCCAGCTACTACAAAGACAACA 954
 Qy 305 AlaIleGlyThrPhe-----ArgAlaMetGluArgGlyIleSerValIle 319
 Db 955 TTCTTGATCATCTTGTGCTACACTCTTGGGCTGTAAAAGCTGTGTTTGTGCT 1014
 Qy 320 CysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAsnThrAlaProTrpVal 339
 Db 1015 CAAGCTGCTGAAACGGAGGTCCTTTCCGAAAACCTCTGGTTTCATACAGCCCTTGGATA 1074
 Qy 340 SerThrIleGlyAlaGlyThrLeuAspArgArgPheProAlaValValArgLeuAlaAsn 359
 Db 1075 ACTACTGTGGTGTGCAATGATGATCGCAGATACAAAATCATCTGACTCTTGGAAAT 1134
 Qy 360 GlyLysLeuLeuTyrGlyGluSerLeuTyrProGlyLysGlyIleLysAsnAlaGlyArg 379
 Db 1135 GGTAAATGCTTGGCGGAATAGGATATCTCTTCTACTCGACCTCATCGTTCATACAG 1194
 Qy 380 GluValGluValIleTyrValThrGlyLysAspLysGlySerGluPhe----- 395
 Db 1195 ATGGTTTCTGCAATGATGTTCTGTTGGTCTTCTGTGTAATAATCAATCCGTCGGAC 1254
 Qy 396 CysLeuArgGlySer---LeuProArgGluGluIleArgGlyLysMetValIleCysAsp 414
 Db 1255 TGCCAGAAAGCCAGAGTCTTGAACAAAAATAGTTCGAAGGAAACATCTCTGTTTGGGA 1314
 Qy 415 ArgGlyValAsn-----GlyArgSerGluLysGlyGluAlaValLys 428
 Db 1315 TATTTCTTCAATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1374
 Qy 429 GluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsn-GlnGluGluAspSe 448
 Db 1375 CATCTAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1434
 Qy 448 rIleAspValHisLeu---LeuProAlaThrLeuIle---GlyTyrThrGluSerValLeu 466
 Db 1435 GATCCCGCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1494
 Qy 467 LeuLysAlaTyrValAsnAlaThrVal----- 475
 Db 1495 TTGATTGATTACTACAAATGTCAACACGTCAAGAGATTGGATGGGAAGGGTAAAGGACTTT 1554

Qy 476 LysProLysAlaArgIlellePheGlyGlyThrValIleGlyArgSerArgAlaProGlu 495
Db : : : : :
1355 AAAGCTGAAGAGACGCGAGACGGTTTGGAAACCCATTCTTCAAAATCCGACCTGAA 1614
Qy 496 ValAlaGlnPheSerAlaArgGlyPro-----SerLeuAlaAsnProSer 510
Db : : : : :
1615 GTTCTCTGTCTCAGCTCGAGGACCAATACCAAGATTTCAGCTTTCAGGATGCTGAT 1674
Qy 511 IleLeuLysProAspMetIleAlaProGlyValAsnIlelleAlaLaTppProGlnAsn 530
Db : : : : :
1675 CTTCTCAACACAGATATCTTGTCTCAGGCTTTAATATAGTCTGCTGCTGCTGCAAT 1734
Qy 531 LeuGlyProThrGlyLeuProTyroAspSerArgValAsnPheThrValMetSerGly 550
Db : : : : :
1735 ---GGAACAGACGAGGCTAACTATATCGCGAA-----GGAATTGCACTAAATTTCTGGC 1785
Qy 551 ThrSerMetSerCysProHisValSerGlyIleThrAlaLeulleArgSerAlaTyroPro 570
Db : : : : :
1786 ACAAGCATGGTGCACACACATTCGCGGTATTGCTCTGTGTGAACAGAGACATCT 1845
Qy 571 AsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThrAlaAspLeuTyroAspArg 590
Db : : : : :
1846 CAATGGAGTCCAGCTGCATTAATACAGCTTTGTATGACAGCTTCAACAGTCATAGATAGA 1905
Qy 591 GlnGlyLysAlaIlelleLysAspGlyAsn-----LysPro 601
Db : : : : :
1906 GCAGGAGACCTCTCCAGACACACAATATCTGAACAGACAGACAGTAACACTTGTATA 1965
Qy 602 AlaGlyValPheAlaIlelleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGly 621
Db : : : : :
1966 GCAACTCGCTTGTATTATGAAGCGGTATGTCAATCAAGCGGTCTGTAGACCTGGT 2025
Qy 622 LeuValTyroAsnIleGlnProValAspTyroIleThrTyroCysThrLeu---GlyPhe 640
Db : : : : :
2026 CTCATCTTGTATGAGGCTTATGAGGATATATAGGTTCTTGTGCACACACACCTGGTATC 2085
Qy 641 ThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuArg 660
Db : : : : :
2086 GATGCTCAGGATGATAAAACTTCACA-----AACACTCCATGCAATTTCAAAATGGTT 2139
Qy 661 LysAsnProGlyPheSerLeuAsnTyroSerIleAlaValIlePheLysArgGlyLys 680
Db : : : : :
2140 CATCTTCAAACTTC-----AACACGCCATCCATGCCATC-----TCTCATCTCGTC 2187
Qy 681 ThrThrGluMetIleThrArgArgValThrAsnValGlySerProAsnSerIleTyroSer 700
Db : : : : :
2188 AGAACACAAACCGTAACACAGAGAGTACGAAATTTTCGGAAGAGAGAGAAACATACACA 2247
Qy 701 ValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPhe 720
Db : : : : :
2248 ATCATCTGAGGATGAGCCAGCAATCGCATCGAAGTGAGTCTCTCTGCAATGACAGTA 2307
Qy 721 Lys---HisValAspGlnThrLeuSer----- 728
Db : : : : :
2308 AGACGGCGGCTTCTAGACCTTTTCGGTGACTCTACAGTACAGTACAGTACGAGCT 2367
Qy 729 TyrArgValTrpPheValLeuLysLysLysAsnArgGlyGlyLysValAla-----Ser 746
Db : : : : :
2368 TATAGCTTCGACAGGTTACATTGAAGGAGCGGCGGATTAAGTGACTCTCTCTGCTG 2427
Qy 747 PheAlaGlnGlyGln 751
Db : : : : :
2428 GTTCTATGGGACAA 2442

RESULT 15

US-10-260-238-901
; Sequence 901, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Krepis, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Rieke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 901
; LENGTH: 2018
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N region
; LOCATION: (442)..(442)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-901

Alignment Scores:
Pred. No.: 9,5e-80 Length: 2018
Score: 793.00 Matches: 248
Percent Similarity: 47.42% Conservative: 129
Best Local Similarity: 31.19% Mismatches: 254
Query Match: 19.74% Indels: 165
DB: 12 Gaps: 27

US-09-806-767-2 (1-775) x US-10-260-238-901 (1-2018)

Qy 10 llellePheLeuPheCysSerSerSerSerGluIleLeuGlnLysGlnThrTyrlle 29
Db : : : : :
6 ATTATATCTTCTTCTCGTGCATTTCTACTACTTACTTACTTTTGCAC-----ATTGTT 56
Qy 30 ValGln-----LeuHisProAsnSerGluThrAlaLysThr--- 41
Db : : : : :
57 GTTCAGACACACGGATCTCGTAAGCTTTACATAGCATATCTCGGTGATGTAACACAG 116
Qy 42 -----PheAlaSerLysPheAspTrpHisLeuSerPheLeuGln 54
Db : : : : :
117 CCACCCCGACGATGTCATCGCTTCACACCATGATATGCTCACCACGTGTTCTTGG----- 170
Qy 55 GluAlaValLeuGlyValGluGluGluGluGluGluGluGluGluGluGluGluGlu 74
Db : : : : :
171 -----AAGAGCAAGGAAGATTCTTTGGGCTCTTATCATCCACAAC 209
Qy 75 TyrGlySerAlaIlelleGlyPheAlaAlaGlnLeuThrGluSerGluAlaGlu-Ilele 94
Db : : : : :
210 TACAAGCATGTTCTCAGGCTTCGCCGTGATGTTACTGAAGATCAAGCAAGCAACTT 269
Qy 94 uArgTyroSerProGluValValAlaValArgProAspHisValLeuGlnValGlnThrTh 114
Db : : : : :
270 GCAGGAATTCGGGAAGTCTTAGTGTGGAACCAAGCAAAACGTATATACGAGCGGACAC 329
Qy 114 rTyroSerTyroLysPheLeuGlyLeuAspGlyPheGlyAsnSerGlyValTrpSerLysSe 134
Db : : : : :
330 TCGGAGCTGGGACATGCTTGGACTGAACATACAGATGCCCACCTGAACCTCCAGCGAAC 389
Qy 134 rArgPheGlyGlnGlyThrIlelleGlyValLeuAspThr-GlyValTrpProGluSerP 154
Db : : : : :
390 CAACCTACGGCGAGGAGATAATCATTTGGCATTTGTTACACTGGGAGTC----- 436
Qy 154 roSerPheAspAspThrGlyMetProSerIleProArgLysTrpLysGlyIleCysGlnG 174
Db 437 -----TGCCANG 443

QY 174 luGlyGluSerPheSerSerCysAsnArgLysLeuIleGlyAlaAaPhePheI 194
Db 444 TCGCGCAGGATGGCGCAGCAAAAT---GAGCGGCAAGATATCGCGCGCGTTA--- 496
QY 194 leArgGlyHisArgValAlaAsnSerProGluGluSerProAsnMetProArgGluTyrI 214
Db 497 -----CAGCGCGGGTGGACGAGGACGACCTCAAGATCGACTACC 536
QY 214 leSerAlaArgAspSerThrGlyHisGlyThrHisThrAlaSerThrValGlyGlySerS 234
Db 537 TCTCGCCCGCGCAGCGGCGCATGGCACGCGCACACGCGGTCACCGCGCGCGGCTCGG 596
QY 234 erValSerMetAlaAsnValLeuGlyAsnGlyAlaGlyValAlaAaArgGlyMetAlaProG 254
Db 597 TCGTGGAG-----GCGGGGTCCGGGAG----- 619
QY 254 lyAlaHisIleAlaValTyrLysValCysTrpPheAsnGlyCysTyrSerSerAspIleL 274
Db 620 -----ACGCCACCGCTGC 632
QY 274 euAlaAlaIleAspValAlaIleGlnAspLysValAspValLeuSerLeuGlyG 294
Db 633 TGGCGCGCATCGACGACGCGATCCACGACGCGGTGGACGCTGCTGCTCCTCGGAGA 692
QY 294 lyPheProIleProLeuTyrAspAspThrIleAlaIleGlyThrPheArgAlaMetGluA 314
Db 693 AC-----TCCTTCGCGCGCGACGACGCGCGCTCCAGA 722
QY 314 rgGlyIleSerValIleCysAlaAlaGlyAsnAsnGlyProIleGluSerSerValAlaAa 334
Db 723 AGGGA---TCACCGTGTACGCGCCATGAACCTTGGCCCTGCGCCTCAGTGGTTCAGA 779
QY 334 snThrAlaProTrpValSerThrIleGlyAlaGlyThrLeuAspArgAaGpPheProAlav 354
Db 780 ACATCGCCCTTGGTCTATCACCGTCGCGGCGACGACATCGATCGCTCTCCGACGG 839
QY 354 alValArgLeuAlaAsnGlyLysLeuLeuTyrGlyGluSerLeuTyr---ProGlyLysG 373
Db 840 TGATCATGTTGGGTGACAGCAGACAGATAGTGGACAACTCTCTACTCTCAAGGGAAGA 899
QY 373 lyIleLysAsnAlaGlyArgGluValGluValIleTyrValThrGlyGlyAspLysGlys 393
Db 900 ACTCATCGCTCAGCGGTTTCAGACGCTTGTCTGTGGAGTT-----GGAG 944
QY 393 erGluPheCysLeuArgGlySerLeuProArgGlu-GluIleArgGlyLysMetValIle 412
Db 945 CTGAATGG-----CACGGATGTCGAAGGACGATTTGCTG 980
QY 413 Cys-----AspArgGlyValAsnGlyArgSerGluLysGlyGluAla 426
Db 981 TGGCGATCTTTTACTCTCAACAAACCATCTCTCTTTCCAAAGAAGCTCTAGGAACGTC 1040
QY 427 ValLysGluAlaGlyGlyValAlaMetIleLeuAlaAsnThrGluIleAsnGlnGluGlu 446
Db 1041 GTGAAA---GGCGGAGGGGTGGCGATTTTCTTCAGTATACGTGGGATATTGAAGT 1097
QY 447 AspSerIleAspValHisLeuLeuProAlaThrLeuIleGly---TyrThrGluSerVal 465
Db 1098 TCGACAGCCAGATCAACCGGTATTTCGTGTGTATTTCGACTATTACAGATGAACACAG 1157
QY 466 LeuLeuLysAlaTyrValAsnAlaThrValLysProLysAlaArgIleIlePheGlyGly 485
Db 1158 ATCGGAAATACATCTACTGTCGAGAGTTCCACCATTTGAGATCGATCCA-GCAGCG 1216
QY 486 ThrValIleGlyArgSer---ArgAlaProGluValAlaGlnPheSerAlaArgGlyPro 504
Db 1217 ACCGTAAAGGGGAATGAGATAATGGCACCACAAAGTGGCAGATTCTCTTCAAGAGGCCCT 1276
QY 505 SerLeuAlaAsnProSerIleLeuLysProAspMetIleAlaProGlyValAsnIleIle 524
Db 1277 TCAACTGATTACCTGAAATATTCAAGCCAGACATAGCTGCACCGAGATTCAACATCTTA 1336

QY 525 AlaAlaTrpProGlnAsnLeuGlyProThrGlyLeuProTyrAspSerArgValAsn 544
Db 1337 GCAGCA-----GTGAAGGCACA 1354
QY 545 PheThrValMetSerGlyThrSerMetSerCysProHisValSerGlyIleThrAlaLeu 564
Db 1355 TACGCATTGCTCAGGAGCGTCAATGGCTACTCCGCATGTAGCAGCGTGTGTCGCGCTG 1414
QY 565 IleArgSerAlaTyrProAsnTrpSerProAlaAlaIleLysSerAlaLeuMetThrThr 584
Db 1415 CTGAAGCTTGGCACTCAAGCTGTCTACCTGCTGCTCACTAAATCAGCATAGTACCCACA 1474
QY 585 -AlaAspLeuTyrAspArgGlnGlyLysAlaIle---LysAspGly-----AsnLysPr 601
Db 1475 GGCATCTGTACTGATGAGCGTGCATGCCATACTGCGCGAAGGGTTCCTCGGAAGAT 1534
QY 601 oAlaGlyValPheAlaIleGlyAlaGlyHisValAsnProGlnLysAlaIleAsnProGl 621
Db 1535 TGCTGACCATTTGACTATGAGCGGCGGCATCAACCCATAACAGACGAGTGTATCTCTGG 1594
QY 621 yLeuValTyrAsnIleGlnProValAspTyrIleThrTyrLeu---CysThrLeuGlyPh 640
Db 1595 CCTGATTTCACGACATTGATCCACGATTACAAACAAATCTTTGGATGCACCTGTA----- 1649
QY 640 eThrArgSerAspIleLeuAlaIleThrHisLysAsnValSerCysAsnGlyIleLeuAr 660
Db 1650 -----AAGCCATATGTAAGTGCAATCGG-----AC 1675
QY 660 GlyAsnProGlyPheSerLeuAsnTyrProSerIleAlaValIlePheLysArgGlyLy 680
Db 1676 ATCGTGTCTGGTATTATCTGAACCTACCGTCCATCTCAGTCCCGATCTTAGGTATCC 1735
QY 680 sThrThrGluMetIleThrArgValThrAsnValGlySerProAsnSerIleTyrSe 700
Db 1736 GGTC-----GTCGTATCAAGACGCGTCACAAATGTAGCTGAGGTGATGCTGTTTACCA 1789
QY 700 rValAsnValLysAlaProGluGlyIleLysValIleValAsnProLysArgLeuValPh 720
Db 1790 CGCTGCAATCGAGAGCCACCTGCGTAAGATGGATGTTGAGCCACCTGTTCTGGTGT 1849
QY 720 eLysHisValAspGlnThrLeuSerTyrArgValTrpPheValLeuLysLysLysAsnAr 740
Db 1850 CAATGCTGCAACAAAGTTTCATAGTCCAGTTAAGCTTTCACCTCTATGGAAGCTGCA 1909
QY 740 gGlyGlyLysValAlaSerPheAlaGlnGlyGlnLeuThrTrpValAsnSerHisAsnLe 760
Db 1910 GGGG-----GACTACAGTGTGGAGCCTTACTTGG-----CACAAAGG 1948
QY 760 uMetGlnArgValArgSerProIleSerValThrLeu 772
Db 1949 CCAGAAAGACAGTGAAGATCCCGATAGCAGTCAAGGATT 1985

Search completed: February 3, 2004, 06:34:23
Job time : 1710 secs